Optimal Estimation Retrieval of Aerosol Microphysical Properties in the Lower Stratosphere from SAGE II Satellite Observations

A thesis
submitted in partial fulfillment
of the requirements for the Degree of
Doctor of Philosophy
in Physics
by

Daniela Wurl

Department of Physics and Astronomy
University of Canterbury
2007

Abstract

A new retrieval algorithm has been developed based on the Optimal Estimation (OE) approach, which retrieves lognormal aerosol size distribution parameters from multiwavelength aerosol extinction data, as measured by the Stratospheric Aerosol and Gas Experiment (SAGE) II in the lower stratosphere. Retrieving these aerosol properties becomes increasingly more difficult under aerosol background conditions, when tiny particles ($\ll$ 0.1 μm) prevail, to which the experiment is nearly or entirely insensitive. A successful retrieval algorithm must then be able (a) to fill the 'blind spot' with suitable information about the practically invisible particles, and (b) to identify 'the best' of many possible solutions. The OE approach differs from other previously used aerosol retrieval techniques by taking a statistical approach to the multiple solution problem, in which the entire range of possible solutions are considered (including the smallest particles) and characterized by probability density functions.

The three main parts of this thesis are (1) the development of the new OE retrieval algorithm, (2) the validation of this algorithm on the basis of synthetic extinction data, and (3) application of the new algorithm to SAGE II measurements of stratospheric background aerosol. The validation results indicate that the new method is able to retrieve the particle size of typical background aerosols reasonably well, and that the retrieved uncertainties are a good estimate of the true errors. The derived surface area densities (A), and volume densities (V) tend to be closer to the correct solutions than the directly retrieved number density (N), median radius (R), and lognormal distribution width (S).
Aerosol properties as retrieved from SAGE II measurements (recorded in 1999) are observed to be close to correlative *in situ* data. In many cases the OE and *in situ* data agree within the (OE and/or the *in situ*) uncertainties. The retrieved error estimates are of the order of 69% (σ_{N}), 33% (σ_{R}), 14% (σ_{S}), 23% (σ_{A}), 12% (σ_{V}), and 13% (σ_{Reff}). The OE number densities are generally larger, and the OE median particle sizes are generally smaller than those N and R retrieved by Bingen *et al.* (2004a), who suggest that their results underestimate (N) or overestimate (R) correlative *in situ* data due to the 'small particle problem'. The OE surface area estimates are generally closer to correlative *in situ* profiles (courtesy of T. Deshler, University of Wyoming), and larger than Principal Component Analysis (PCA) retrieval solutions of A (courtesy of L. W. Thomason, NASA LaRC) that have been observed to underestimate correlative *in situ* data by 40-50%.
These observations suggest that the new OE retrieval algorithm is a successful approach to the aerosol retrieval problem, which is able to add to the current knowledge by improving current estimates of aerosol properties in the lower stratosphere under low aerosol loading conditions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Stratospheric aerosol is important for a number of processes that affect the chemical and radiative balance of the atmosphere and is known to be able to significantly perturb the climate system in the aftermath of large volcanic eruptions. Investigating aerosol particle sizes and abundance with the help of satellites is therefore an important area of climate research. The retrieval of those aerosol microphysical properties is, however, not straightforward and sufficiently complex, so that even after more than 30 years of research a number of important questions remain unanswered, that determine the focus of ongoing and future research [SPARC, 2006]. One of those areas of interest is the exact determination of the microphysical properties of very small stratospheric particles that have not been volcanically influenced. These so-called *background aerosols* are the focus of the research documented in this work.
This chapter starts by providing an overview of how atmospheric aerosols influence the Earth's climate (Sec. 1.1). The difficulties in accurately retrieving aerosol microphysical properties are introduced in Section 1.2; Section 1.3 presents the the main objectives of this work and an outline of the thesis content.

1.1 Atmospheric Context: Aerosols and Climate

An aerosol is defined in its simplest form as a collection of solid or liquid particles suspended in a gas. Common usage however refers to the aerosol as the particulate component only. In the atmosphere, aerosol particles exhibit a large variety of chemical composition, sizes and shapes, and cause a wide range of phenomena such as dust, fume, smoke, mist, fog, haze, clouds, and smog. In the lower parts of the atmosphere, particularly in the planetary boundary layer, aerosols can strongly affect visibility, human health and our quality of life [Hinds, 1998; Colbeck, 1998]. These phenomena occur with time scales ranging from several hours to days because of the generally well mixed nature of the troposphere.
In the upper troposphere and stratosphere, aerosols generally have much longer lifetimes

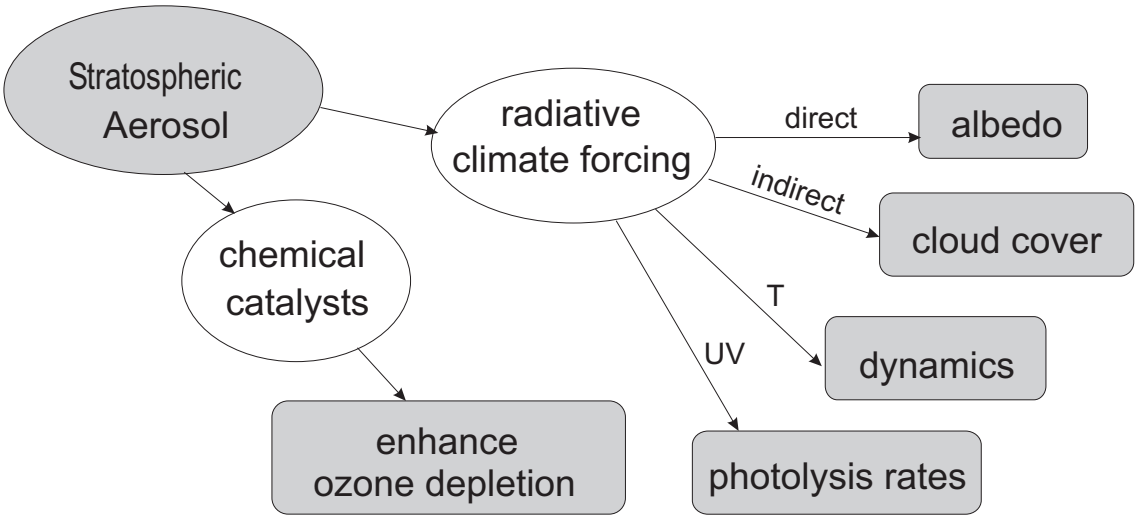

Figure 1.1: Aerosol effects on local and global climate.

(one to two years for stratospheric aerosols, [Seinfeld and Pandis, 1998]) and can therefore affect the Earth's climate on a regional to global scale, even though they are much less abundant than in the troposphere. Figure 1.1 provides an overview of the major mechanisms and ways that stratospheric aerosols influence physical as well as chemical processes in the atmosphere.

- **Radiative impact**

  During volcanically quiescent periods, when the stratospheric aerosol is in a 'background' state unperturbed by volcanism, radiative effects of the aerosol are negligible [SPARC, 2006]. However, following a major eruption the increased aerosol loading can have a transient but significant radiative impact [McCormick *et al., 1995*]. Direct effects involve scattering and absorption by particulates. Stratospheric aerosol particles with strong scattering properties like sulphuric acid (H_2SO_4) droplets decrease the total energy of the Earth-atmosphere system directly by scattering a large portion of the incoming solar radiation immediately back into space. This increases the albedo of the atmosphere which leads to an associated cooling effect. Such effects are usually quantified in terms of the radiative forcing, which is the net flux change at the top of the atmosphere due solely to the direct aerosol radiative effects. Pueschel (1996) estimates that after the eruption of Mount Pinatubo in June 1991 the solar radiation reaching the Earth's surface and troposphere by September 1991 changed by as much as - 2.7 Wm^{-2}. Although there are uncertainties in the estimates of aerosol radiative forcing, it is generally agreed that the averaged global direct effects of anthropogenic sulphate aerosols can locally be similar in magnitude but in opposite direction to anthropogenic greenhouse gas forcing.

Volcanic aerosols may also have indirect effects on the troposphere and the total Earth-atmosphere energy budget, when the particles act as condensation nuclei and change the microphysical structure and hence the optical properties, extent and lifetime, of clouds. A growing number of small cloud droplets increase the cloud reflectivity and therefore indirectly increases their albedo. Since on the other hand clouds reduce the loss of terrestrial radiation to space, thereby enhancing the atmospheric greenhouse effect, aerosols can be responsible for partially offsetting the albedo effect [Kiehl et al., 2000]. Comparing the radiative forcing by volcanic aerosols with the effect of the CO_2 build-up in the atmosphere since the Industrial Revolution, which in 1998 was estimated to have caused a net change of about $+1.25\,\mathrm{Wm}^{-2}$ [Hinds, 1998], it becomes clear that stratospheric aerosols can have an effect on global surface temperature, that is similar in magnitude but in the opposite direction to climate warming by greenhouse gases. A change in temperature in turn will have an influence on the dynamical processes in the atmosphere.

- **Atmospheric Dynamics**

  Some aerosol types, like soot and volcanic ash, are strong absorbers of solar radiation. This local heating of the stratosphere induces changes in its radiative balance, which in turn results in a dynamic feedback. Model simulations suggest that the pole-to-equator temperature gradient observed after the eruption of Mount Pinatubo [Robock and Mao, 1995] led to an enhanced polar vortex and a wave response in the tropospheric circulation [Graf et al., 1993; Kodera, 1994; Timmreck et al., 1999].

- **Chemical Processes**

  The chemical balance of the stratosphere can be influenced by a changed radiation field, in particular in the ultraviolet range of the electromagnetic spectrum. This leads, for instance, to changes in ozone photolysis rates, which in turn influences the budget of several trace gases, in particular NO_x [SPARC, 2006].Alternatively, aerosol particles can act as chemical catalysts serving as sites for heterogeneous reactions. Ozone depletion can then be strongly promoted, for example through changing the partitioning of nitrogen species and converting chlorine from inactive into active forms, and by changing the frequency of occurrence of Polar Stratospheric Clouds (PSCs) [Granier and Brasseur, 1992]. Heterogeneous processes occurring on the surface of ice particles in PSCs are known to play a key role in the formation of the springtime ozone hole over Antarctica [Solomon et al., 1986; Hofmann and Solomon, 1989].

The above examples give an idea of the intricate interactions involving aerosol particles, radiation processes, chemistry and dynamics in the atmosphere. The magnitudes of the different forcing mechanisms demonstrate the importance of stratospheric aerosols in climate

research. Quantitative knowledge of aerosol abundance and their extinction coefficients is also required for the analysis of spectroscopic climate data. If not properly accounted for, the presence of aerosols can adversely affect the retrieval results of other parameters, for example sea surface temperatures or trace gas mixing ratios. This in turn will affect the estimation of climate trends.
Since virtually all properties of atmospheric aerosols depend strongly on the particle size distribution and associated surface area and volume densities, the accurate determination of those aerosol microphysical properties is essential in aerosol science and climate studies. Although important knowledge about aerosol properties can be and has been gained from *in situ* measurements by probing air samples (e.g. Junge and Manson, 1961, or more recently by Deshler *et al.*, 1992 and thereafter), global coverage can only be achieved by using satellite measurements. Evaluation of such measurements with respect to the aerosol microphysical properties requires the use of retrieval techniques, since satellite instruments measure aerosols remotely and indirectly by recording the effect of the particle population on radiation.

1.2 The Retrieval Problem

Generally, all measurements that are a function of a particular quantity of interest represent a retrieval problem. The cause-effect sequence which gives rise to the measurements is called a *direct* or *forward* problem. Aerosol particles, for instance, interact with electromagnetic radiation and give rise to characteristic extinction features in the solar spectrum, which can be measured by a satellite instrument. The mathematical reversal of the cause-effect sequence, which consists of finding the unknown causes (e.g. aerosol properties) of known consequences (e.g. measured extinction) is the associated *inverse* or *retrieval* problem. Retrieval problems can be very complex where the measurement of several different quantities are used to infer several other quantities which describe the state of some system. This is the case for the problem at hand where the size distribution of aerosol particles is determined from multispectral aerosol extinction as measured by the space-based Stratospheric Aerosol and Gas Experiment (SAGE) II.

All practical problems have in common that in the presence of measurement uncertainty retrieval solutions are not unique. Retrieval is therefore first of all a matter of identifying all physically acceptable solutions which reproduce the data to within the experimental error. As the set of mathematically acceptable solutions can be very broad and encompass completely different answers, all physically sensible answers (a retrieved length, for instance, must be positive) have to be selected before one solution can be identified as 'the best' solution. Ideally, the best solution is the one which is closest to the true solution. However, since the true solution is naturally unknown, the final answer will be

one which best fulfills some criteria. For some situations it is also possible that no retrieval solution may exist at all. This is the case where the information contained in the measurements is not sufficient to retrieve the number of unknowns. Such problems are called 'under-constrained' or 'ill-posed'. Prior information can, in the form of so-called *a priori* constraints, provide the additional information necessary to solve the equations and identify a set of possible solutions. Then, one of those solutions can be selected as the 'best' solution.

When evaluating space based measurements of aerosol extinction aerosol scientists are faced with two more problems. One concerns the difficulty of detecting particles smaller than approximately 0.1 μm through available remote sensing techniques. At SAGE II wavelengths, for example, the contribution to the total extinction from particles smaller 0.1 μm (in a sample lognormal particle size distribution with a median radius of 0.07 μm) is on the order of 15% at 0.386 μm and 5% at 1.020 μm. In comparison, SAGE II extinction errors are typically between 10 and 40% at 0.386 μm, and 7% or smaller at 1.020 μm, which is the same magnitude as the contributions to the total extinction from the smallest particles. This means that contributions by those very small particles, typically found in volcanically quiescent periods, may lie well within the experimental error. (In theory, higher sensitivity can be achieved by shortening the observational wavelength. In practice, however, ozone absorption in the lower stratosphere imposes an effective lower wavelength limit of $\sim$0.38 μm for solar occultation measurements of aerosol.) In addition, Steele *et al.* (1997) estimate that particles with radii between 0.1 and 1.0 μm can contribute to 76% of the surface area and to 90% of the total volume, and simultaneously represent 30% of all the particles. From these results they conclude that errors arising from tiny particles (smaller than approximately 0.1 μm) affect the surface area and volume much less than the size distribution parameters [Steele and Turco, 1997].

The second fundamental problem is encountered when discriminating different particle sizes. This problem arises from the linear dependence between Mie scattering cross sections at different wavelengths. Application of the 'ratio-criterion' [Heintzenberg *et al., 1981*] to SAGE II measurements, which analyses ratios of effective cross sections at different wavelengths as a function of particle radius, suggests effective upper and lower limits in particle radius of about 0.1 μm and 1.0 μm for size discrimination [Wang *et al., 1989*]. At smaller radii, the signal becomes very small, and at larger radii, the dependence on particle radius strongly decreases as the extinction efficiency approaches the value 2.

As a result of both the low sensitivity and the size discrimination problems, aerosol properties that are heavily dependent on either tail of the size distribution, like number density and median particle radius, are more error prone [Twomey, 1977]. This probably explains why the lower order moments are still poorly constrained, whereas more can be found in the literature on integrated aerosol properties such as surface area density, volume density and effective radius.

In the past decade significant effort has been devoted to obtaining a better understanding of aerosol properties, and a number of different retrieval approaches have been applied to the aerosol retrieval problem. Recent observations and data comparisons show that during low aerosol loading periods such as the current period, the operational SAGE II retrieval algorithm tends to underestimate surface area densities as derived from *in situ* measurements by optical particle counters [e.g. Deshler *et al.*, 2003; SPARC, 2006]. Since it turns out that aerosol surface area density is controlled primarily by particles which contribute minimally to extinction [Steele *et al.*, 1999; SPARC, 2006], it has been suggested that a possible reason for underestimating A could be that the operational retrieval model puts little material in this 'blind spot' that results from the low sensitivity to the smallest particles at visible wavelengths [SPARC, 2006]. For the same reason the retrieved volume density can be expected to be underestimated, and concomitantly the effective radius overestimated by the SAGE II retrieval algorithm. Based on the disagreement between the various data sets and models, the authors conclude that significant questions remain regarding the ability to characterize stratospheric aerosol during volcanically quiescent periods, particularly in the lower stratosphere [SPARC, 2006].

In summary, when retrieving aerosol size distribution parameters from SAGE II measurements under low aerosol loading conditions we are mainly faced with the following difficulties:

- The aerosol retrieval problem is ill-posed, (as will be shown), such that the particle size distribution cannot be retrieved without the use of additional constraints.
- The lack of a unique solution requires the identification of the 'best' of many possible solutions, according to some criteria.
- Very small aerosol particles (<0.1 μm), that are typical for stratospheric background conditions, are hard to detect as their contribution to the total extinction is of the same order of magnitude as the measurements uncertainty.

As a result:

- Aerosol properties that heavily rely on those particles, like aerosol number density and median particle radius, are particularly hard to retrieve.
- These lower order moments are still poorly constrained.
- The retrieved aerosol properties are very sensitive to the model used in the retrieval algorithm and dependent on model assumptions regarding the effectively invisible small aerosol particles.

1.3 Objectives and Thesis Outline

In order to address the open questions named above a retrieval algorithm was developed which is based on a novel approach to the aerosol retrieval problem and which has the following characteristics:

1. In contrast to other methods used so far, the presented Optimal Estimation (OE) algorithm takes a statistical approach to solve the multiple solution problem. This approach is comprehensive in that the entire range of possible solutions are considered and characterized by probability density functions.

2. The solutions are constrained by assuming lognormal particle size distribution and by using prior knowledge about the solution space which is used in the form of a probability density functions.

3. The 'blind spot' which arises from the low sensitivity of solar occultation measurements (at visible wavelengths) to small particles is addressed by allowing for particle radii smaller than 0.1 μm. Including particles smaller than 0.1 μm is - if done appropriately - not only important for the low order moments of the size distribution (number density and median radius), but also for higher order moments such as surface area density.

4. The OE retrieval solution comprises the most likely value as well as an associated uncertainty estimate.

The new retrieval algorithm is tested on synthetic spectral extinction data and then applied to SAGE II measurements recorded in 1999 with the following objectives:

- gain new knowledge of aerosol particle size distributions, which so far are rather poorly constrained;

- to improve current estimates of aerosol surface area density and volume density under aerosol background conditions;

- to apply the new results to study seasonal differences of monthly mean latitude-altitude cross sections of aerosol properties under background conditions;

and thus to contribute to the current knowledge of aerosol microphysical properties in the lower stratosphere under low aerosol loading conditions.

This thesis is structured as follows:

Chapter 2 provides the theoretical basis and the scientific context of this work. The aerosol microphysical properties and a common mathematical description of particle ensembles are introduced. The relevant radiative transfer equations and the physics of aerosol extinction are presented. We describe the Stratospheric Aerosol and Gas Experiment (SAGE) II satellite observations, to which the new algorithm will be applied, introduce currently used retrieval methods, and describe the results obtained from these algorithms.

Chapter 3 presents the development of the new algorithm. The mathematical building blocks are introduced and the model equations are developed. The information content of the SAGE II spectral channels is assessed and the retrieval problem characterized by analysing the sensitivity of the new algorithm to various parameters. Based on this information, the OE algorithm is refined and tailored to the specific problem of retrieving aerosol size distribution parameters from the four spectral SAGE II channels. The chapter closes with an analysis of the various sources of retrieval uncertainty.

Chapter 4 contains an internal validation of the new algorithm. Based on two sets of synthetic aerosol extinction measurements (same aerosols but two different measurement noise scenarios), the accuracy of the retrieved aerosol properties is assessed by direct comparison with the correct solutions. A quality filter is developed that allows the automatic identification of bad results. The retrieved uncertainties are evaluated and additional possible sources of error assessed. Finally, by looking at the altitude structure of the retrieved and true aerosol properties, possible reasons for the observed discrepancies are identified.

Chapter 5 presents the results achieved by applying the new Optimal Estimation retrieval algorithm to real data, namely to SAGE II measurements of aerosol extinction. The retrieved aerosol properties and associated uncertainties are described and compared to aerosol properties as derived from coincident *in situ* measurements. The OE integrated properties are then compared to the SAGE II operational retrievals of surface area density and effective radius (obtained through Principal Component Analysis, PCA). Finally, the new retrieval results are used to generate monthly zonal averages (latitude-altitude cross sections) of various aerosol properties. These are used to study seasonal differences in background aerosol properties on a near global scale during the year 1999.

Chapter 6 closes the work with a summary of the main results, conclusions and areas of future work.

Chapter 2

Theoretical Background

During the past decade, considerable effort has been devoted to obtaining a better understanding and characterization of stratospheric aerosols. The development and refinement of a variety of different instrumental techniques and measuring geometries now allows us to measure the effects of stratospheric aerosols on large temporal and spatial scale, and those measurements are evaluated by various research teams using different retrieval approaches. This chapter provides some information about the microphysical properties of atmospheric aerosols, how they can be measured, and which methods are used to retrieve aerosol properties from indirect measurements. Section 2.1 introduces a commonly used mathematical model to describe the size distribution of an ensemble of particles and provides mathematical expressions for integrated quantities such as the surface area and volume densities. Section 2.2 illustrates how aerosols influence radiation as it propagates through the atmosphere. The basic mathematical equations as well as some typical aerosol extinction spectra are also presented. Section 2.3 gives an overview of satellite experiments measuring aerosols, then introduces in more detail the SAGE II instrument and data products used in this study. Section 2.4 briefly introduces retrieval approaches that have been or are being used by other research teams. Some of the results derived using these retrieval techniques and their associated uncertainties are also described.

2.1 Atmospheric Aerosol Properties

The microphysical properties of aerosols comprise the particle size distribution and the associated integrated properties (surface area and volume). These properties are largely determined by the source or production mechanisms of the particles, subsequent chemical reactions, growth and removal processes. Source material which enters the atmosphere already as particles forms *primary aerosols*, whereas *secondary aerosols* are formed in the atmosphere by gas-to-particle conversion. Primary aerosol particles can already be large when they enter the atmosphere, as for example soot particles. Secondary aerosols, for

instance sulphuric acid droplets that are typically found in the lower stratosphere, start as tiny particles that grow with time according to particle dynamics and the available supply of material. Measurements with *in situ* and remote sensing instruments have shown that the majority of all stratospheric aerosol particles reside between the tropopause and 30 km, that is in the lower part of the stratosphere. In a volcanically perturbed atmosphere a particle maximum can be observed between 18 and 22 km [Lu *et al., 1997*].

Composition

Generally, the most common types of aerosols found in the Earth's atmosphere are non-absorbing sulphate aerosols, organic carbon aerosols, desert dust aerosols, and after volcanic eruptions volcanic ash aerosols. In the troposphere, the large number of different aerosol species can be grouped into sulphur-, nitrogen-, carbon-, and halogen-containing compounds [Seinfeld and Pandis, 1998]. The composition of aerosols in the troposphere is far more diverse than in the stratosphere, where the non-volcanic natural stratospheric aerosol (or 'background' aerosol) at low and mid-latitudes, is typically composed of an aqueous sulphuric acid solution [Junge and Manson, 1961]. At temperatures between 193 and 218 K (-80 to -45°C) the sulphuric acid concentration (acidity) is typically between 60 and 80% by weight [Shen et al., 1995]. The natural background level of stratospheric sulphuric acid aerosol particles mainly originates from photochemical modification of the sulphur-containing biogenic carbonyl sulphide (OCS) gas [Turco *et al.*, 1980] which has its sources at the Earth's surface. Because of its low reactivity in the troposphere and its correspondingly long residence time, OCS is the most abundant sulphur gas in the global background atmosphere, and it is the only tropospheric sulfur component that survives to enter the stratosphere [Seinfeld and Pandis, 1998]. In the stratosphere, OCS is photodissociated and attacked by oxygen atoms and OH radicals, and the gaseous product of the chemical breakdown of OCS is SO_2. This SO_2 is subsequently converted to H_2SO_4 aerosol. The background loading of sulphuric acid aerosol is estimated to amount to approximately 20 kT and typical particles are about 0.05 μm in size [Hofmann, 1990].

Major volcanic eruptions can increase the particle size by a factor of ten and the sulphate loading by several orders of magnitude. For example, Mount Pinatubo (1991) ejected an estimated 20 MT of sulphur dioxide (SO_2) and a total of 30 MT of material into the stratosphere [Bluth et al., 1992]. The SO_2 gas is converted over a period of months into H_2SO_4-H_2O droplets [Bluth *et al.*, 1992; Grant *et al.*, 1996], whereas the removal of the resulting submicron-size aerosol particles from the stratosphere by sedimentation processes takes several years [Hofmann, 1990]. This makes volcanoes the dominant source of enhancement and variability in the stratospheric aerosol loading.

Anthropogenic contributions to stratospheric aerosols include high altitude aircraft emissions of SO_2 and soot [Pollack et al., 1976; Turco et al., 1980; Hofmann, 1990], space

shuttle Al_2O_3 particulate emissions [Hofmann *et al.*, 1975] and industrial emissions of OCS (transported through the tropopause into the stratosphere).

The average chemical composition of aerosols in the polar stratosphere is different from stratospheric aerosols at mid-latitudes. Although at temperatures greater than approximately 196 K and altitudes near 20 km sulphate particles are expected to be present, colder temperatures enable the formation of polar stratospheric clouds (PSC). This means that H_2SO_4-H_2O droplets sequester nitric acid (HNO_3) to form ternary H_2SO_4-HNO_3-H_2O solution particles with a large variety of ratios depending on the ambient stratospheric conditions. As the temperature approaches the ice frost point (~ 188 K) a considerable amount of nitric acid and water vapour condenses on the sulphate aerosol, and the fraction of H_2SO_4 in the ternary particle becomes very small [Molina *et al.*, 1993]. A more detailed description of PCSs, their formation processes and classification can be found in Seinfeld and Pandis (1989).
Near the tropopause, where tropospheric and stratospheric air mix, other elements can be detected. Especially during intense convective activity over populated areas, particles at high altitude may contain diverse material such as nitrate, ammonia, organics, minerals, and metals [e.g. Heintzenberg *et al.*, 1996; Talbot *et al.*, 1998; Chen *et al.*, 1998; Murphy *et al.*, 1998]. However, the detailed chemical composition of particles in the tropopause region is only poorly characterised [Kärcher and Solomon, 1999].
In the present study, the stratospheric aerosol is assumed to be composed of sulphuric acid droplets. There is ample evidence to support this assumption [e.g. Rosen, 1971; Russel *et al.*, 1981a; Oberbeck *et al.*, 1983].

Particle Size Distributions

In the atmosphere, many different particle sizes coexist. Measurements of stratospheric aerosols by *in situ* particle counters [Deshler *et al.*, 1993] and thermodynamical calculations have shown that aerosol particle populations cover about five orders of magnitude in size. The size spectrum is generally continuous and may range from only a few nanometres up to several hundred micrometres. A range of different statistical models have been fitted to the size distributions of different aerosol types. Among these are the power law [e.g. Thomason, 1991], the exponential distribution [e.g. Junge and Manson, 1961], a modified gamma distribution [e.g. Wang *et al.*, 1989] and the lognormal distribution [e.g. Russel *et al.*, 1981; Deshler *et al.*, 1993]. Descriptions and comparative presentations can be found in Deepak and Box (1982) or Hinds (1998). The most widely used size distribution model for stratospheric aerosols is the differential lognormal expression

$$\frac{dN(r)}{d\ln r} = \frac{N}{\sqrt{2\pi}\,S} \cdot \exp\left[-\frac{1}{2}\frac{(\ln r - \ln R)^2}{S^2}\right], \tag{2.1}$$

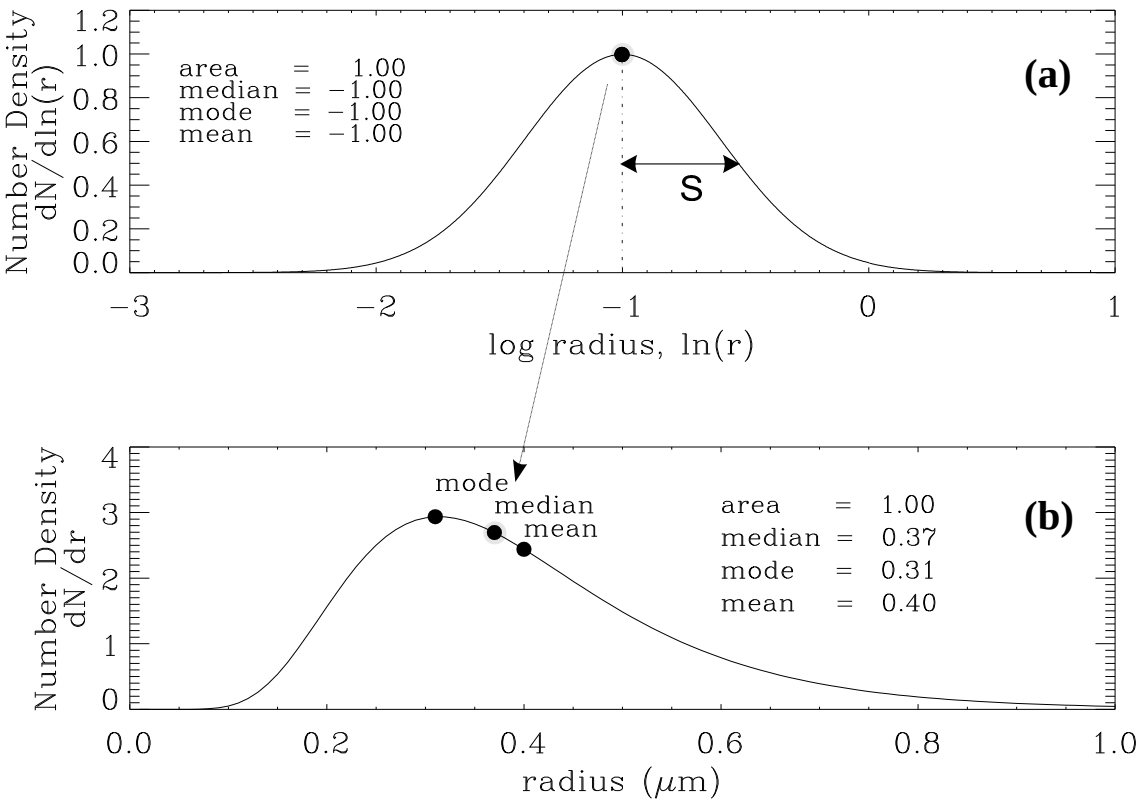

Figure 2.1: A log-normal distribution in logarithmic and linear coordinates.

where N is the total number of particles per unit volume of air, R is the median radius, and S is the spread or half width (mostly just called 'width') or standard deviation of the Gaussian distribution. $dN/d\ln r$ describes the number of particles per unit volume of air per logarithmic particle size increment, that is all particles in a radius interval between $\ln r$ and $\ln r + d\ln r$. The total number of particles (usually in cm^{-3}) can be calculated by summation over all particle radii:

$$N = \int_{-\infty}^{\infty} \frac{N(r)}{d\ln r} d\ln r \quad = \quad \int_{0}^{\infty} \frac{N(r)}{dr} dr. \tag{2.2}$$

For practical reasons it is often convenient to know the frequency distribution in linear space, i.e. as a function of r rather than $\ln r$:

$$\frac{dN(r)}{dr} = \frac{N}{\sqrt{2\pi}\,S} \cdot \frac{1}{r} \cdot \exp\left[-\frac{1}{2}\frac{(\ln r - \ln R)^2}{S^2}\right]. \tag{2.3}$$

Compared to the distribution in log space n(r) is not a symmetric Gaussian but skewed towards smaller radii. The relationship between the log-normal function and its transformation into linear space is illustrated in Figure 2.1. Conversion from logarithmic into

linear space shows that the centre value of r of the log-normal distribution is identical with the median radius of the function in linear radius space. The mode radius R_{mod} and the arithmetic mean, R_{mean}, of the particle size distribution are related to the median radius R and spread S by

$$\ln R_{\mathrm{mod}} = \ln R - S^2 \tag{2.4}$$

$$\ln R_{\mathrm{mean}} = \ln R + \frac{1}{2}S^2. \tag{2.5}$$

It is worthwhile noting that in the literature values of $\sigma = \exp S$ are often given rather than S.

The monomodal log-normal size distribution has been found to apply to most single-source aerosols [Hinds, 1998] and generally describes non-volcanic background conditions very well. In the volcanically disturbed atmosphere, superposition of different particle modes results in multimodal particle size distributions. Typically, distinct bimodal and sometimes even trimodal size distributions can be detected [e.g. Massie *et al.*, 1996]. These distributions are simply expressed by a superposition of several lognormal distribution functions:

$$\frac{dN(r)}{dr} = \sum_i \frac{N_i}{\sqrt{2\pi}\,S_i} \cdot \frac{1}{r} \cdot \exp\left[-\frac{1}{2}\frac{(\ln r - \ln R_i)^2}{S_i^2}\right]. \tag{2.6}$$

The primary mode, which is the smallest in terms of particle size, is generally referred to as the *nucleation mode*. As the primary mode particles grow, e.g. by coagulation and condensation they form a secondary mode called the *accumulation mode*. In the stratosphere this mode develops after volcanic eruptions and is highly effective for extinction of light. Together, the nucleation and accumulation modes contain the *fine particles*. The third mode is the *coarse particle mode* and consists of large particles that are generated by mechanical processes and are directly emitted into the atmosphere. This mode causes extremely high extinction of light [Grainger et al., 1993]. The size intervals associated with the three modes are (in radius) approximately 0.002-0.05 μm (nuclei mode), 0.05-1.2 μm (accumulation mode), and larger than 1.2 μm (coarse mode) [Seinfeld and Pandis, 1998].

Bimodal distributions can also occur in the upper troposphere, where the few available field observations indicate that high number densities ($> 10^3 \mathrm{cm}^{-3}$) of quite small particles ($< 0.3\,\mu$m) can be present superimposed on a mode of larger but less abundant particles. Kärcher and Solomon (1999) suggest that such a bimodal spectrum may develop in situations where new particle formation and growth occur in the presence of a pre-existing aerosol.

Integrated Properties

The aerosol surface area density and volume density can be calculated based on the *moments* of a size distribution. These are quantities that are proportional to the particle size raised to a power (moment). The j$^{\text{th}}$ moment, M_j, is given by:

$$M_j = \int_0^\infty r^j \cdot \frac{dN(r)}{dr}\, dr. \tag{2.7}$$

This is the j$^{\text{th}}$ moment about zero, which should not be confused with moments about the mean, which define the variance and skewness of a distribution. From Equation 2.7 it follows that $M_0 = N$, the zero order moment is equal to the particle number density. For the lognormal monomodal and multimodal size distributions, Equation 2.7 can be shown to be:

$$M_j = \sum_k N_k R_k^j \cdot \exp\left[\frac{j^2}{2} S_k^2\right], \tag{2.8}$$

where $k = 0$ for a single mode and $k = 0, 1, \ldots n$ for n modes. Consequently, the surface area and volume densities of monomodal aerosols are given by

$$A = \int_0^\infty 4\pi r^2 \cdot \frac{dN(r)}{dr}\, dr = 4\pi M_2 = 4\pi N R^2 \cdot \exp\left[2S^2\right] \tag{2.9}$$

$$V = \int_0^\infty \frac{4}{3}\pi r^3 \cdot \frac{dN(r)}{dr}\, dr = \frac{4}{3}\pi M_3 = \frac{4}{3}\pi N R^3 \cdot \exp\left[\frac{9}{2}\, S^2\right]. \tag{2.10}$$

The surface area density is usually given in $\mu\text{m}^2\text{cm}^{-3}$, and the volume density in $\mu\text{m}^3\text{cm}^{-3}$. From A and V the effective radius or area-weighted mean radius of a size distribution can then be derived. It is

$$R_{\text{eff}} = \frac{M_3}{M_2}. = \frac{3V}{A} = R \cdot \exp\left[\frac{5}{2} S^2\right] \tag{2.11}$$

The total particle surface area per unit volume of air, A, and the particle volume, V, (or equivalently mass) per unit volume of air play a key role in determining the chemical and radiative impact of aerosols on climate. The surface area density is an important parameter to the photochemistry community since numerous studies have suggested that enhancements of aerosol surface areas due to volcanic eruptions may lead to heterogeneous reactions similar to those that occur on polar stratospheric clouds but on a smaller scale [Solomon et al., 1986]. Where the reacting species have time to penetrate the surface and become mixed the particle volume becomes important for the chemical reactions. The volume density can also be used to determine the mass concentration and mixing ratio.
The effective radius is an indicator of the radiative impact of the particles. Investigations of climate forcing by stratospheric aerosol led Lacis *et al.* (1992) to the conclusion that the effective radius is indicative of the aerosol climate forcing. According to their observations an effective radius greater than about 2 μm indicates that the global average greenhouse

effect of the aerosols exceeds the albedo effect, resulting in a net surface heating. Effective radii smaller than approximately 2 μm indicate cooling.
Since aerosols are largely variable in space and time, the observed values are usually a function of altitude and vicinity (in time and space) to volcanic eruptions. Under background conditions, aerosol surface area densities (as derived from *in situ* measurements at the University of Wyoming, see Sec. 4.3) are typically of the order of 0.1-1.0 $\mu m^2 cm^{-3}$, volume densities 0.01-0.10 $\mu m^3 cm^{-3}$, and effective radii on the order of 0.1 μm. Under volcanic influence, surface area densities as large as 50 $\mu m^2 cm^{-3}$ [Lambert *et al., 1997*] or 35 $\mu m^2 cm^{-3}$ [Grainger et al., 1995] have been observed, the latter with an associated volume density of 6 $\mu m^3 cm^{-3}$ and an effective radius of 0.5 μm [Grainger et al., 1995].

2.2 Radiative Transfer and Aerosol Extinction

The presence of atmospheric aerosols can be detected based on their effect on other processes in the atmosphere, for instance on the propagation of sunlight. The sum of scattering and absorption of electromagnetic radiation by aerosols is called *aerosol extinction*.
The intensity, I, of electromagnetic radiation transmitted through an inhomogeneous medium decreases with increasing distance, s, measured along the propagation path. This decrease is observed to be exponential and can be described by the Beer-Lambert law:

$$I = I_0 \exp\left[-\beta^{\text{ext}} \cdot s\right], \tag{2.12}$$

where I_0 is the initial intensity, and β^{ext} the *volume extinction coefficient* at a particular wavelength. The dimensionless product $\beta^{\text{ext}} \cdot s$ is the *optical depth* and describes the total extinction along the line-of-sight. The extinction properties of a medium depend on the efficiency with which it it removes light out of the beam by absorption and scattering. The volume extinction coefficient can be thought of as the cross-sectional area per unit volume with which the ray interacts. It is the sum of all particle cross-sections multiplied by an efficiency factor

$$\beta^{ext}(\lambda) = \pi \int_0^\infty r^2 \cdot Q^{ext} \cdot \frac{dN(r)}{dr} dr \quad = \quad \int_0^\infty C(r,\lambda) \cdot \frac{dN(r)}{dr} dr, \tag{2.13}$$

where $dN(r)/dr$ describes the number of particles at size (radius) r, Q^{ext} is the *extinction efficiency*, which depends on the particle size, the wavelength of the incident light, and the refractive index, and $C(r,\lambda)$ is the (number) kernel function. The extinction coefficient is conventionally given in $\mu m^2 cm^{-3}$ or km^{-1} and often just called 'extinction'. For simplicity we have also adopted this nomenclature and will use the word extinction as a short form for 'volume extinction coefficient'. Expressed in terms of particle volume Eq. 2.13 is

$$\beta^{ext}(\lambda) = \int_0^\infty \frac{3Q^{ext}(r,\lambda)}{4r} \cdot \frac{dV}{dr} dr \quad = \quad \int_0^\infty \widetilde{C}(r,\lambda) \cdot \frac{dV}{dr} dr, \tag{2.14}$$

where dV/dr is the volume density of particles per unit radius interval and $\widetilde{C}(r, \lambda)$ stands for the (volume) kernel function. Both C and $\widetilde{C}$ are important functions because the efficacy of any technique for retrieving aerosol attributes is limited by the kernels of the integral equation from which extinction is derived, and the kernel functions are largely determined by the extinction efficiency factor.

Mie Scattering Theory

The need for an extinction efficiency factor arises from the nature of light extinction by small particles. Scattering and absorption of light involves the interaction between electromagnetic waves and the electric charges that constitute matter. When a beam of light impinges on a particle, electric charges in the particle are excited into oscillatory motion. They act as an array of oscillating multipoles that give rise to secondary electric and magnetic waves, which combine in the far field to produce a scattered wave. Light scattering means that energy is reradiated in all directions. Absorption takes place when the incident radiation is converted into thermal energy. The efficiency of scattering and absorption is determined by the *size parameter*, which is defined as the ratio of particle circumference to the wavelength,

$$x = \frac{2\pi\, r}{\lambda}, \tag{2.15}$$

and by the complex refractive index of the aerosol. If a particle is much smaller than the incident wavelength ($x \ll 1$), for instance air molecules in visible light, a particle acts as a single dipole. The resultant scattering is known as *Rayleigh scattering*. Particles for which $x \gg 1$ fall into the so-called geometric scattering regime. This applies for instance to cirrus cloud particles and aerosols forming in the polar winter Arctic and Antarctic vortices. In this case the scattering can be determined on the basis of the geometrical optics of reflection, refraction, and diffraction. Scattering is then strongly dependent on particle shape and orientation relative to the incoming beam. At particle sizes similar to the size of the wavelength ($x \approx 1$) the scattering pattern becomes very complex as the particle appears as an array of multipoles. The scattered radiation at any given point is then taken as the superposition of the secondary waves generated by the multipoles (which constitute the particle) as they oscillate in the incident electric field.

Gustav Mie (1908) first solved Maxwell's equations for scattering on homogeneous spheres. In this case the angular distribution of light intensity can be expressed by an infinite series of Ricatti-Bessel functions and Legendre polynomials. His work is known as *Mie scattering theory*. Mie theory has shown that when a particle disturbs the radiation field an electromagnetic wave can be affected beyond the geometric extent of the particle. It provides an expression for the extinction efficiency, which is the ratio of the effective and the geometric extinction cross-sections. Van de Hulst (1957) discusses Mie theory in detail

substance	H_2SO_4	$(NH_4)_2$
temperature (K)	220-350	250-350
wavelengths (nm)	250-1100	240-800
mass fraction (%)	3-100	5-50

Table 2.1: Validity ranges of the new refractive index model developed by Semmler *et al.* (2003) for binary $(NH_4)_2$-H_2O or H_2SO_4-H_2O and for ternary $(NH_4)_2$-H_2SO_4-H_2O solutions.

together with many of its applications. For sulphate aerosol models the assumption that the droplets are spherical is justified since surface tension forces liquid sulphuric acid particles resulting from the gas-to-particle transformation process to produce spherical shapes [Torres et al., 1998]. The forward model which was developed as part of this research is based on a Mie scattering code that originates from the work of Grainger (1990) and is available at www.atm.ox.ac.uk/code/mie.

Aerosol Refractive Index

Generally, the refractive index of a substance is a measure of the weakening and retardation of a propagating wave in this medium. It is a complex number and can be described by

$$m = n - i\,k = \sqrt{\epsilon_0 - \frac{2\kappa\lambda}{c}i}, \tag{2.16}$$

where ϵ_0 is the dielectrical constant, κ the electric conductivity of the particle, and c is the speed of light in vacuum [van de Hulst, 1957]. κ is a function of the composition of the particle, which in the case of sulphuric acid is determined by the ambient temperature and the water vapour partial pressure.

The real part, n, is the customary refractive index, which gives the ratio of the speed of light in vacuum to the speed of light in a medium (absolute index) or between two different materials (relative index), for instance between air and an aerosol droplet. As the refractive index of air is effectively 1 ($m = 1.00029 - 0.00000$i at λ =589 nm), the relative index of the particle can in practise be taken as the absolute index. The imaginary part, k, is a damping factor and determines the decrease of light intensity by absorption. Since the different ions present in an aerosol have different spectral characteristics, the refractive index changes with the wavelength of the incident light.

At 1.06 μm the refractive index of sulphuric acid droplets ranges between 1.394 and 1.444 [Steele *et al.*, 1999]. This range covers ambient conditions typically found in the lower stratosphere with temperatures between 195 K and 240 K, water vapour pressures of $1 \cdot 10^{-4}$ to $8 \cdot 10^{-4}$ hPa, and associated acidities between 35 and 85% H_2SO_4 by

weight. Slightly different values may arise for ammonium sulphate particles or other impurities, which can be found in the upper troposphere and occasionally also in the lower stratosphere. The refractive index of ammonium sulphate aerosol is approximately $m = 1.520-0.000i$ at 0.55 μm, and $m = 1.510-0.000i$ at 1.02 μm [Rosen, 1991]. These example refractive indices indicate that both ammonium sulphate and sulphuric acid solution aerosols are strong scatterers and that absorption is practically zero at visible wavelengths.

At spectral wavelengths where the imaginary part of the refractive index is zero ($m = n$), which is the case for sulphuric acid in the visible, the refractive index at cold temperatures can be calculated from the refractive index at a reference temperature. The dependence of the refractive index $n(T)$ on droplet density $\rho(T)$, as described by Yue *et al.* (1994), is given by

$$n(T) = \sqrt{\frac{2L \cdot \frac{\rho(T)}{\rho(T_0)} + 1}{1 - L \cdot \frac{\rho(T)}{\rho(T_0)}}}, \tag{2.17}$$

where T_0 is a reference temperature at which n and ρ are known, and L is the Lorentz-Lorenz correction factor

$$L = \frac{n(T_0)^2 - 1}{n(T_0)^2 + 2} = \frac{4\pi}{3}\frac{\alpha}{\rho(T_0)}, \tag{2.18}$$

where α is the specific polarizability. With the help of the Lorentz-Lorenz relationship the refractive index at a certain temperature can be calculated from the complex refractive index at a reference temperature. For the correction, α is assumed to be independent of temperature, and the particle density ρ must be known at both reference and target temperatures. In order to assess the validity of the correction for sulphuric acid Pinkley and Williams (1976) applied the Lorentz-Lorenz relationship to refractive indices measured at room temperature (by Palmer and Williams, 1975) and compared the transformed values with measurements of m (for 75% and 95.6% H_2SO_4) between 1.67 μm and 25 μm at 250 K. They concluded that the correction worked well on the real part of the refractive index far from absorption bands, but that there were significant differences at the band maxima for the imaginary part of the refractive index.

For retrieving aerosol properties from SAGE II satellite measurements (Ch. 5) the refractive index is determined using a model by Semmler *et al.* (2003) which, based on new laboratory measurements of the densities and refractive indices of binary or ternary H_2SO_4, and/or $(NH_4)_2SO_4$ and water solutions, calculates refractive indices for a wide range of upper tropospheric/lower stratospheric temperatures and acidities (mass fractions) using the Lorentz-Lorenz relationship. The validity ranges of the model are given in Table 2.1. The uncertainties associated with indices of binary H_2SO_4-H_2O solutions are estimated to be smaller than 2% (Semmler, personal communication, 2003).

As at the time of the model validation neither the above refractive index model nor laboratory measurements of sulphuric acid refractive indices at stratospheric temperatures and visible wavelengths were available, the OE algorithm was evaluated using refractive indices measured at room temperature originating from the work of Palmer and Williams (1975). The use of room temperature data here is not a problem because the same refractive indices were used to generate the synthetic spectra and to retrieve the aerosol parameters from these spectra.

Aerosol Concentration

The refractive index of an aerosol particle is a function of its composition. The composition of a sulphuric acid aerosol particle, which is also called concentration or acidity, is defined as the mass fraction of H_2SO_4 and water. It is determined by the thermodynamics of an H_2SO_4-H_2O drop in equilibrium. An expression relating the equilibrium drop size and composition to the ambient water vapour partial pressure and temperature was derived by Nair and Vohra (1975). Further investigations of changes in temperature and humidity on the growth and optical properties of aerosol drops were made by Steele and Hamill (1981) and Russell and Hamill (1984). They pointed out that the equilibrium with water vapour is so fast that over those short time scales any changes in the H_2SO_4 vapour pressure can be neglected. This means that once the sulphuric acid has condensed (equilibrium state) the equilibrium particle size is determined by the ambient water vapour and temperature, and the aerosol acidity is basically independent of the mass of acid in the drop. Lambert *et al.* (1996) showed that this makes the H_2SO_4 component of the H_2SO_4-H_2O aerosol a conserved tracer.

When evaluating SAGE II measurements, the aerosol acidity was determined with the help of the observed humidity and atmospheric pressure (SAGE II data) and the prevailing temperature (National Meteorological Center, NMC) data. The humidity is given as a volume mixing ratio

$$\mathrm{vmr} = V_\mathrm{e}/V_\mathrm{d}, \tag{2.19}$$

that is the ratio of the volumes of water vapour (V_e) and dry air (V_d). In combination with the ideal gas equation

$$p_\mathrm{k} V_\mathrm{k} = nRT_\mathrm{k} \tag{2.20}$$

where n is the number of moles, R the universal gas constant ($R = 8.3145$ J/(mol K)), p_k is the partial pressure, V_k the volume and T_k the temperature of a particular substance k, and expressing the total pressure as a sum of all partial pressures (dry air and water vapour)

$$p = p_\mathrm{d} + p_\mathrm{e} \tag{2.21}$$

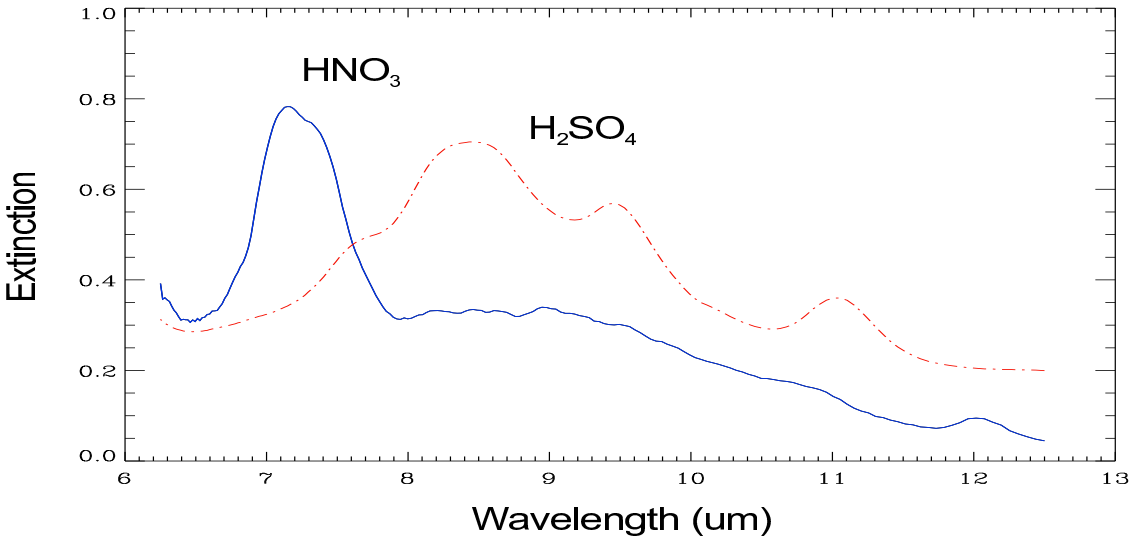

Figure 2.2: Optical depth spectra of 70% HNO_3-H_2O and H_2SO_4-H_2O binary aerosols.

then the water vapour partial pressure can be determined by

$$p_e = \frac{p}{\frac{1}{\text{vmr}}\frac{M_e}{M_d} + 1} \tag{2.22}$$

with the molar masses of water vapour and dry air being $M_e = 18.02$ g/mol and $M_d = 28.96$ g/mol. With temperature and water vapour partial pressure, the H_2SO_4 weight fraction in H_2SO_4-H_2O aerosol concentration can be determined by linearly interpolating between tabulated values from Steele and Hamill (1981) and extensions from Russell and Hamill (1984).

Spectral Information

The extinction spectrum of a particular aerosol draws its characteristic shape from the chemical composition and temperature of the substance as well as from the size distribution of the particle population. The chemical composition determines the locations of extinction minima and maxima through the molecular structure of a particle. At the molecular level, the chemical bonds determine the quantum mechanical energy states and associated possible energy transitions, which, in turn, give rise to specific spectral signatures. The energy transitions depend above all upon the strength of the atomic and intermolecular bonds and the wavelength or wavenumber of the incident electromagnetic wave. A detailed description of the physics of spectroscopy can be found in Atkins (1997).

Figure 2.2 depicts two example spectra (optical depth) of a sulphuric acid aerosol and a nitric acid aerosol. It can be observed that the relative maxima and minima of the two substances are at distinctly different locations. It can also be observed that both

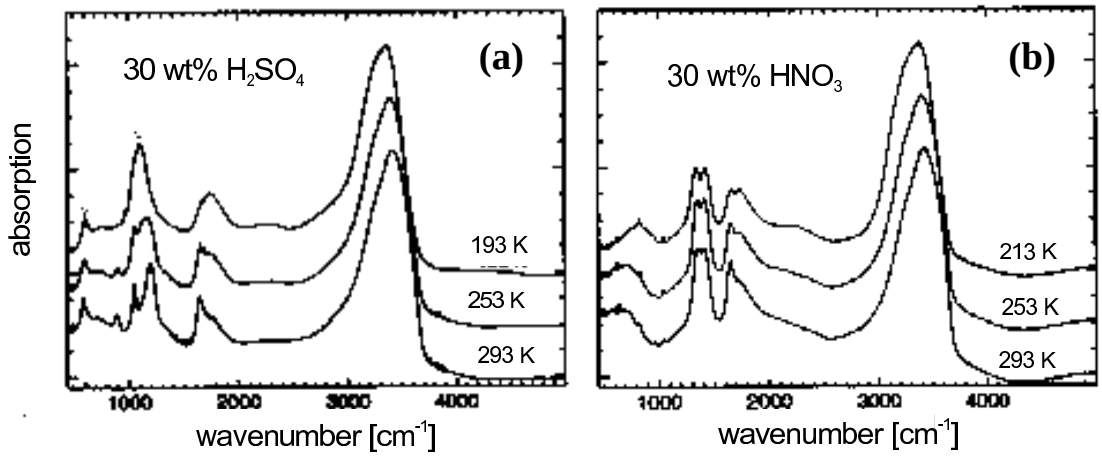

Figure 2.3: Absorption spectra of 30% H_2SO_4-H_2O (a) and HNO_3-H_2O (b) aerosols at different temperatures [Biermann et al., 2000].

substances show broadband absorption features, which are characteristic of liquid and solid particles. They are different from the distinct line spectra, which are generated by gaseous emission or absorption.

The temperature, acidity and particle size distribution of an aerosol determine the intermolecular forces and hence the possible energy transitions (absorption) and the angular redistribution of an incident electromagnetic wave (scattering). Figure 2.3 and Figure 2.4 (taken from Biermann *et al.*, 2000) show the effects of temperature and composition (acidity) on the absorption features of sulphuric acid and nitric acid binary solutions in the mid-infrared spectral region. Investigations by Biermann *et al.* (2000) have shown that at those wavelengths (between approximately 2 and 20 μm) the dependence of sulphuric and nitric acids on temperature and concentration is particularly pronounced.

With respect to scattering and absorption, the sulphuric acid aerosol extinction spectrum can generally be divided into two main radiation regimes. At wavelengths smaller than 2.5 μm the imaginary part of the complex refractive index is approximately zero such that aerosol droplets act as scatterers and the term extinction can be used as a synonym for scattering. At wavelengths longer than 2.5 μm the ratio of absorption to scattering increases rapidly and absorption becomes the major contributor to the total extinction. In terms of the size parameter, an aerosol can be considered to act purely as an absorber if the size parameter decreases to less than about 0.25 [Grainger et al., 1995], that is if the particle is much smaller than the wavelength ($r \ll \lambda$). Grainger *et al.* (1995) have also observed that at wavelengths greater than about 3 μm the absorption is weakly dependent on the shape of the drop size distribution and is approximately a linear function of the

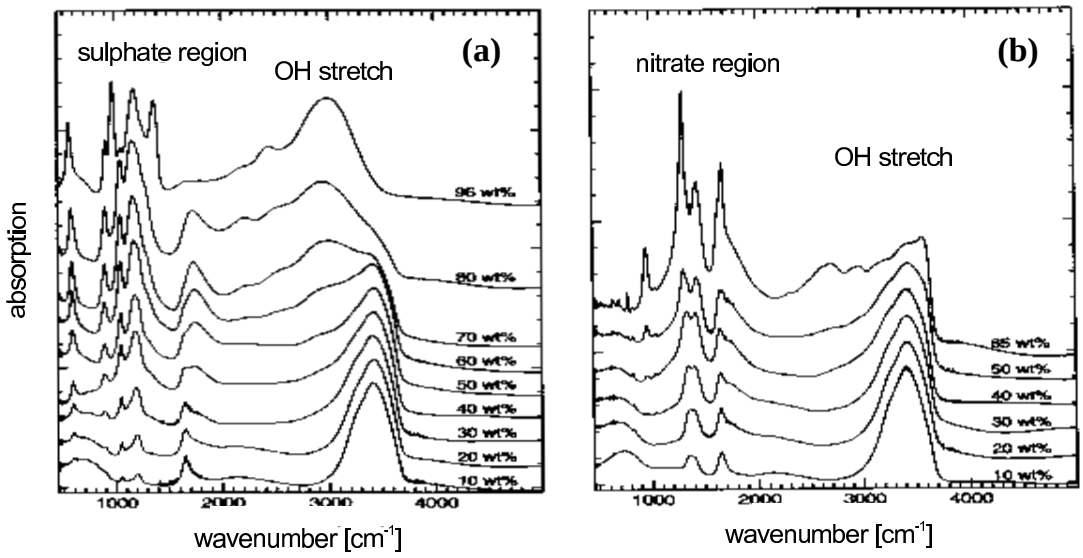

Figure 2.4: Dependence of the spectral absorption characteristics of sulphuric acid (a) and nitric acid (b) (at 293 K) on acidity, from Biermann *et al.* (2000).

aerosol volume [Grainger et al., 1995]. This can be explained by the observation that in the absorption regime the efficiency factor Q^{abs} is proportional to the size parameter $2\pi r/\lambda$, which means that the volume kernel function $3Q_{\text{ext}}/4r$ (Eq. 2.14) is approximately constant.

As for the retrieval of particle size distributions the sensitivity of the spectral extinction data to particle size is essential, stratospheric aerosol size distributions are best retrieved from aerosol extinction (scattering) spectra measured at optical wavelengths. This fact was considered in the design of the Stratospheric Aerosol and Gas Experiments (SAGE) which are aimed at investigating stratospheric aerosols (and several trace gases). In order to be able to retrieve the size distribution parameters from a set of spectral extinction measurements each spectral signature must be associated with a particular combination of number density, median radius and distribution width. A description of the spectral dependence of aerosol extinction at SAGE II wavelengths on R and S will be presented in Section 4.1.

2.3 Measuring Aerosols - SAGE II

The presence of aerosol in the atmosphere can be measured either *in situ* by taking local air samples, or remotely with indirect measurement techniques. *In situ* techniques applied from airborne platforms like balloons, aircraft or rockets include backscatter sondes, different kinds of particle collection devices, and optical particle counters. These devices can provide detailed information about the microphysical, chemical and optical properties of aerosol. The geographical range of *in situ* techniques, however, is limited to the immediate vicinity of the instrument. In contrast, remote sensing instruments like lidars and spectrometers – ground based, mounted on aircraft, balloon or space based platforms – can investigate atmospheric properties up to hundreds of kilometres away from the location of the detector. This makes measurements possible in places that are otherwise hard to reach, for instance remote ocean regions. Space based remote sounding instruments are also able to take continuous measurements to systematically cover wide geographical areas. This aspect is particularly valuable for global aerosol studies, as aerosol abundances vary strongly in time and space.

Routine monitoring of the stratosphere started in 1978 with the Stratospheric Aerosol Measurement (SAM) II instrument. This single spectral channel instrument, which was launched on Nimbus 7, was built specifically for monitoring stratospheric aerosol extinction in the 1.0 μm wavelength region. Its extent was restricted to high-latitude regions between 64° to 86° N and S. SAM II stayed operational for 15 years until December 1993.
Another mission, the Stratospheric Aerosol and Gas Experiment (SAGE) I, started in February 1979. SAGE I was a four-channel sun photometer with spectral pass bands centred at 1.02, 0.6, 0.45, and 0.385 μm, the first and third of which where suitable to retrieve aerosol properties. Also, compared to SAM II the geographic reach was much increased, with SAGE I taking measurements nearly globally between 80°N and 80°S. The operational time, however, was not quite three years due to a failure of the electric system. Its successor, SAGE II, is an improved version of the SAGE I experiment. Its great asset is that it provides the longest satellite data set on record to date.

SAGE II

The Stratospheric Aerosol and Gas Experiment II was launched from the space shuttle in October 1984. Mounted aboard the Earth Radiation Budget Satellite (ERBS) it has since then been operational continuously until January 2005.
The SAGE II data represent the longest continuous record of spaceborne measurements of stratospheric aerosol, which makes them particularly valuable for global climate studies. The SAGE II instrument is a seven-channel sun photometer which measures changes in received sunlight as the Sun rises or sets as seen from the spacecraft. Satellite sensors

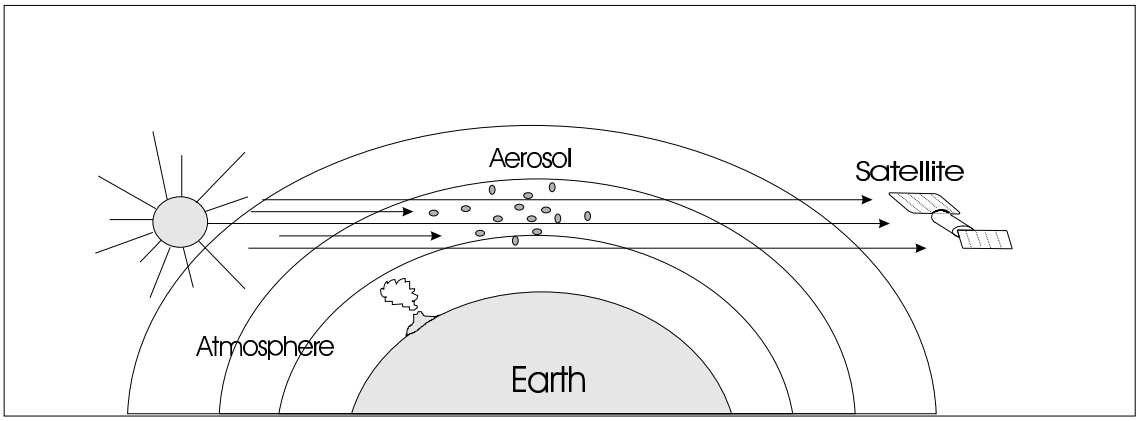

Figure 2.5: Measuring geometry: Solar occultation limb sounding.

that detect the attenuation of the solar signal as it traverses through a portion of the Earth's atmosphere are called *solar occultation* instruments. A schematic illustration of the solar occultation limb-viewing geometry is presented in Figure 2.5. The instrument measures the transmission of the solar irradiance through the limb of the atmosphere. The degree of attenuation of the solar signal depends on the type and abundance of atmospheric components that give rise to light extinction by means of scattering and absorption. A typical SAGE II slant path length is about 200 km long for a 1-km thick shell at a tangent height of about 20 km. The four wavelengths that are used for aerosol retrieval are 1.02, 0.525, 0.452, and 0.386 μm with associated channel bandwidths ranging between approximately 2 ad 20 nm. Each day, SAGE II measures approximately 15 sunrise and 15 sunset events, equally spaced in longitude along two latitude belts between 80°N and 80°S. The latitude location of these measurements during the year depends primarily on spacecraft orbit inclination. Near-global coverage (between extremes of latitudes) is achieved in about a month. For a description of the optical assembly and operation of the SAGE II instrument the interested reader is referred to McCormick (1987) and McMaster (1986). The SAGE II scientific mission objectives as specified by the NASA Langley Research Center (Hampton, VA) are [McCormick, 1987]:

- to map vertical profiles of stratospheric aerosol extinction and ozone, nitrogen dioxide, and water vapour concentrations with vertical resolutions of the order of 1-3 km, and determine high altitude cloud coverage globally from 80°S to 80°N;
- to study the seasonal and global variations in these atmospheric constituents in order to understand their climatology and effects on earth radiation budget and climate;
- to utilise these stratospheric species and tracers to study atmospheric dynamics, sources, sinks and transient phenomena such as volcanic perturbations;

- to investigate atmospheric chemistry involving these species and to study the optical and physical properties of aerosols.

The optical data are recorded at a series of discrete altitudes (tangent heights) for the retrieval of vertical extinction profiles. Before the optical depth for each slant path can be separated into contributions from aerosols and gases, the Rayleigh component (extinction contributions from the background molecular atmosphere) has to be removed first [Chu *et al., 1989*]. A second inversion step is carried out to determine the vertical profile of each species by computing the contribution to the extinction from each atmospheric shell and applying Chahine's (1968) nonlinear inversion technique [Chu and McCormick, 1979]. This way, the measured solar irradiance is numerically inverted to yield vertical profiles of ozone, nitrogen dioxide, water vapour concentration, and aerosol extinction (in km^{-1}). The inversion algorithm of the SAGE II observations is described in articles by Chu (1986) and Chu *et al.* (1989).

The absolute errors on the SAGE II aerosol extinction measurements provided by NASA are $\pm 1\sigma$ uncertainties [Yue et al., 1989]. This means that the distance from a data point to either end of the associated error bar is the expected root-mean-square difference between the SAGE II-derived extinction and the actual extinction. A typical distribution of relative uncertainties associated with the spectral SAGE II measurements is presented in Figure 2.6. It can be observed that on average the uncertainty decreases with increasing wavelength. This is due to a decrease in Rayleigh scattering and gas absorption [Steele and Turco, 1997]. The majority of uncertainties in the 0.386, 0.452, 0.525, and 1.020 μm channels are smaller than 70, 40, 30, and 10%, respectively, with associated most frequent values near 16, 11, 8, and 1%.
The error sources included in the derivation are radiance measurement errors, Rayleigh errors (errors in molecule density arising from uncertainties in the temperature data), reference height errors, and the uncertainty associated with the removal of other species which have overlapping contributions in the spectral wavelength channel. The total error of the inverted extinction at each height level is given by the root-mean-square of these four errors, assuming that they are uncorrelated. Chu *et al.* (1989) describes the four error components as follows:

- The measurement errors in this case are given by the estimated uncertainties from the transmission program in calculating the standard errors of the mean optical depth values at each tangent height level.

- The Rayleigh errors are calculated from the temperature errors given by the NWS (National Weather Service) associated with each temperature profile. The magnitude of the temperature errors generally is within the range from 2°C at sea level to about 12°C at 0.4 mbar pressure level (that is at approximately 55 km).

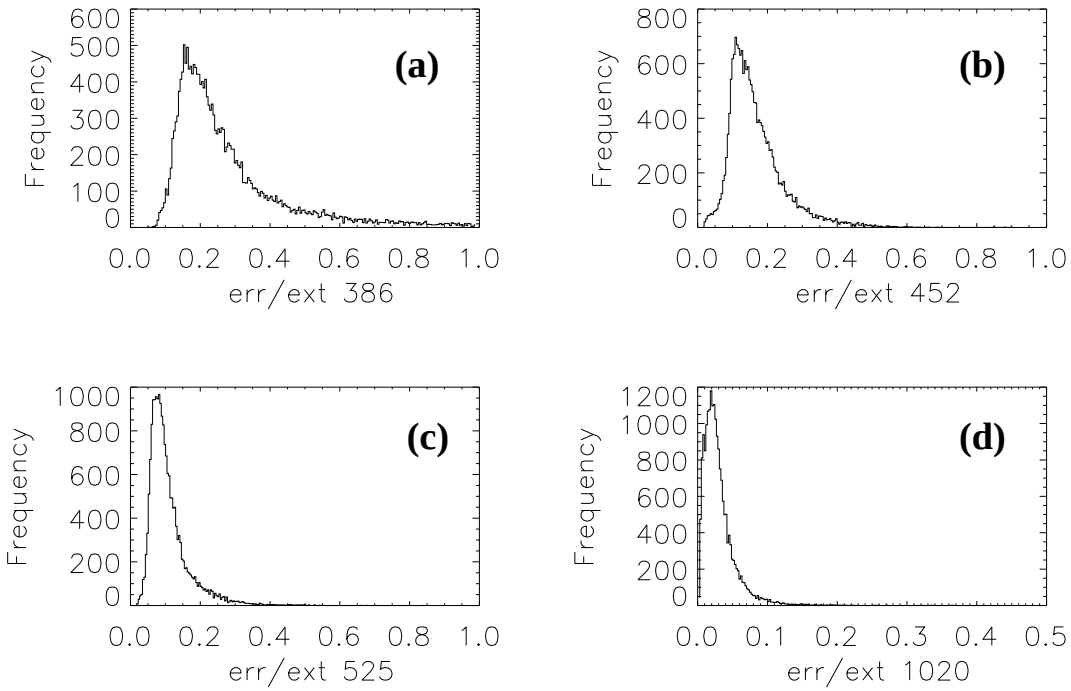

Figure 2.6: Relative uncertainties (ratio of error to extinction) on SAGE II aerosol extinction data (December 1999). The numbers 386, 452, 525 and 1020 refer to the wavelengths (in nm) of the four SAGE II aerosol spectral channels.

- The reference altitude error is the uncertainty in assigning the corrected geometric altitude for each measurement position in order to subtract the Rayleigh components in all of the SAGE II channels. This error is estimated to be about 200 m (1σ) from the spacecraft ephemeris calculations for each SAGE II sunrise or sunset measurement event.

- The errors contributed by the other species arise from the uncertainties in removing contributions from different species in the particular spectral channel.

The SAGE II data which are analysed as part of this thesis are a subset of the version 6.1 data made available to the public by the NASA Langley Research Center (Hampton, VA). A description of this data set is available on the internet at http://www-sage2.larc.nasa.gov/data/v6_data/ and it can be downloaded or ordered from the Langley ASDC at http://eosweb.larc.nasa.gov/PRODOCS/sage2/table_sage2.html.

2.4 Retrieval Methods

The relatively large number of retrieval methods developed and used illustrates the complexity and lack of uniqueness of the aerosol retrieval problem and the continuous attempts to improve the achieved results. Retrieval methods which have been used in the past to investigate aerosol properties include:

- *Principal Component Analysis* (PCA):
  In the PCA approach, the kernel functions in the aerosol extinction equation (Eq.2.14) are expanded in terms of a set of orthogonal basis functions. The idea is that by eliminating the higher-order principal components (basis functions) and truncating the expansion after a few terms, one can exclude from the solution all components for which the noise in the solution, produced by error magnification, exceeds the expected magnitude of the component [Twomey, 1974]. Integral aerosol properties (volume and surface area density) can then be evaluated from a linear combination of the extinction measurements $\beta(\lambda_i)$ multiplied by a factor which is dependent on the expanded kernel function. Further details of this method are described later, as PCA is the method used by the NASA Langley Research Center (LaRC) to retrieve aerosol properties from SAGE II aerosol extinction and also the methodology which the Optimal Estimation inversion results are compared to (Sec. 5.3).

  The PCA retrieval approach has been applied to SAGE II extinction data by a number of investigators to derive aerosol integral properties [Thomason and Osborn (1992), Thomason and Poole (1993) (to Antarctic aerosol), Yue *et al.* (1995), Thomason *et al.* (1997), Steele *et al.* (1999), Lu *et al.* (2003)].

  Steele *et al.* (1999), for instance, investigate systematic errors (arising from the inadequacy of the experiment to measure the desired property) and random errors (resulting from propagation of measurement errors in the recorded extinction) and estimate the trade-off between these two error components by varying the degree of constraint (that is the number of principle components retained). Assuming relative extinction errors of 0.4, 0.4, 0.2, 0.2 at the four SAGE II channels, their study suggests that for lognormal size distributions typical of background aerosols retrieved surface areas can be underestimated by up to 50% through PCA whereas the random error component is estimated to be 15-20% (for the SAGE II instrument).

  Thomason and Poole (1993) estimate an error of approximately 30% in the surface area and 12-35% in the volume, depending on the altitude of the observations.

- *Constrained Linear Inversion* (CLI):
  In the CLI approach [Phillips, 1962; Twomey, 1963; King *et al.*, 1978], solutions are constrained in the retrieval process through the application of a smoothing matrix.

The aerosol extinction integral equation is discretized via quadrature, and the final solution vector contains the particle number at the given size intervals. In order to achieve relatively smoothly varying size distributions, a Junge-type weighting function $h(r)$ can be integrated into the kernel function so that $C(r,\lambda) \rightarrow C(r,\lambda)h(r)$. The inversion can then be carried out for the more slowly varying function $f(r) = dN(r)/dr \cdot h^{-1}(r)$. Assuming that $f(r)$ is constant over each coarse size interval, the extinction equation can be replaced by an equivalent linear system, which can be expressed in the matrix form $g = Af + \epsilon$, where $g_j = \beta_{\lambda_j}$ and β_{λ_j} are the measurements at $j = 1, 2, \cdots J$ discrete wavelengths λ_j, A_{jl} represent the quadrature coefficients $A_{jl} = \int_{r_l}^{r_{l+1}} C(r,\lambda_j)dr$, $f_l = f(r_l \leq r \leq r_{l+1})$ at $l = 1, 2, \cdots, q$ discrete radii r, and ϵ is an unknown error representing the combination between measurement error and quadrature error. The CLI solution equation is then $f = (A^T S^{-1} A + \gamma H)^{-1} A^T S^{-1} g$, where γ is a nonnegative Lagrange multiplier selected so that all the elements of f are positive. Steele and Turco (1997) also transform the radius to a logarithmic scale in order to achieve non-negative physical solutions. Iterations are performed until agreement (in the least squares sense) within the estimated errors is obtained between the computed and the simulated extinction. Anderson *et al.* (2000), for instance terminate an iteration when the root square of the deviation is lower than 18%. CLI is the method employed to extract the aerosol component of extinction within the SAGE II inversion algorithm [Steele and Turco, 1997].

Researchers applying the CLI method to the retrieval of aerosol properties include Anderson *et al.* (2000, applied to simulated SAGE III extinction) and Steele and Turco (1997, applied to SAGE II measurements). Steele and Turco (1997) assessed the effect of experimental error in the integrated properties of the size distributions by carrying out 40 Monte Carlo simulations with a random error at each wavelength of 10%. This was done for several analytic size distribution models. On the basis of the mean and standard deviations and in the presence of 10% extinction error the effective radius could be recovered roughly to within 15%, surface area to within 25%, and volume to within 15% over the particle size range 0.1-1.0 μm in radius. Additional error arises due to particles which lie beyond the size range considered.

- *Randomized Minimization Search Technique* (RMST):
  In this method, developed by Heintzenberg *et al.* (1981), the number size distribution is represented by a histogram whose column heights are varied until a minimum deviation between the corresponding extinction spectra and the experimental data is obtained. The minimization is done by iterative least squares fit. In the beginning of the inversion, each of the size bins is filled with a positive number of particles, one of which is then randomly selected and multiplied by a factor a. The other size bins are multiplied by a factor b. Analytical solutions for the two factors are found by

minimizing the square difference between the associated spectral extinctions and the measurements, that is by setting the first derivative of this difference with respect to a and b equal zero. The iteration step is rejected if a factor becomes negative. The iteration converges and is terminated when the root square of the deviation is lower than a prescribed value within the expected measurement uncertainty of the optical data. (Anderson *et al.* (2000), for instance, allow a maximum deviation from the measurements of 15%). This procedure is repeated a preselected number of times and then the results are averaged to obtain the necessary smoothing, since the individual solutions may be oscillatory in nature (mainly due to the amplification of the measurement noise [Steele and Turco, 1997]).

Investigators who have applied the RMST method to the retrieval of aerosol integral properties from SAGE II extinction data include Lin and Saxena (1992), Saxena (1995) and Anderson and Saxena (1996). Wang *et al.* (1996), Yue *et al.* (1997), and Anderson *et al.* (2000) have applied the RMST approach to simulated SAGE III measurements. Results by Anderson *et al.* (2000), that are based on 10 selected aerosol distribution models, indicate that in the seven-eight SAGE III channel retrievals, both the RMS and CLI techniques obtain total errors in the range 8-50% for the aerosol surface area with an average total error of $\sim$ 25%; estimated total uncertainties of particle volume are in the range 5-25% with an average total error of $\sim$ 12%; for the effective radius the uncertainty range is 6-36% with an average of $\sim$ 20%. They find also that optimal choices for the number of iterations and the number of solutions to be averaged are 100 and 25, respectively. In comparison, the new Optimal Estimation algorithm needs an average of 5 iterations (and no more than 50) and no repeated retrievals since solutions are already smooth by nature.

- *Nonlinear Iterative methods* (NIM):
  In this approach aerosols are assumed to have one of several possible analytical size distributions, which are fitted to observational data by minimizing residual error in some way [Quenzel, 1970; Russel *et al.*, 1981; Yue and Deepak, 1983]. Some investigators have applied NIM to diverse measurements in order to investigate post Pinatubo aerosol [Deshler *et al.*, 1992; Goodman *et al.*, 1994; Russell *et al.*, 1996]. Investigators that applied NIM to the evaluation of SAGE II measurements (with respect to aerosol properties) include Yue *et al.* (1986) and Wang *et al.* (1989).

  Wang *et al.* (1989), for example, adopted a modified Levenberg-Marquardt algorithm for determining model size distribution parameters in the least squares sense and, using both single-mode lognormal and modified gamma representations, achieved uncertainties in median particle radius of between 5% and 28%. The uncertainties of integrated number densities $N_{0.15}$ and $N_{0.25}$ are estimated to smaller than 11%.

In their analysis best retrieval uncertainties were achieved for aerosol radii between 0.1 and 0.7 μm. The error estimates were achieved by using 50 different noise components on each measurement (that were produced by a random noise generator according to the noise level at the SAGE II channels), and calculating the mean and standard deviation. The uncertainties of integrated number densities $N_{0.15}$ and $N_{0.25}$ (that is the number of particles with radii smaller 0.15 and 0.25 μm, respectively) are estimated to be smaller than 11%.

- *Empirical Method* (EM):
  The empirical approach is based on the observation that aerosol surface area and volume densities can be linked by an empirical relationship of the form of a power law $A = \alpha V^{\gamma}$ [Grainger *et al.*, 1995; Lambert *et al.*, 1997; Massie *et al.*, 1998; Steele *et al.*, 1999]. As infra-red aerosol extinctions are strongly correlated with these size distribution moments (extinction is approximately proportional to aerosol volume), these correlations can be used to determine V and A and the associated effective aerosol radius (through Eq. 2.11) in a straightforward manner, provided that the composition and refractive index of the aerosol are known. It has to be considered, however, that the use of the above formulation implies that those *in situ* data from which α and γ were derived are also representative of the aerosol under investigation. This has to be kept in mind as stratospheric aerosol is highly variable in space and time. Investigations by Steele *et al.* (1999) suggest that the empirical formulation leads to altitude dependent systematic errors (due to the difficulty of measuring very small particles) because the relative error depends on the particle number density.

  Hervig *et al.* (1998) expand this method to retrieve particle size distribution parameters as well. They have tabulated effective radii and extinction ratios for a range of aerosol median radii and distribution widths and a particular aerosol composition. The dependence of the effective particle radius on particle mean size and spread is used to compute the dependence of the median radius on the effective radius and extinction ratios, for a range of different aerosol compositions (60-85% in 5% increments). From the measurements, aerosol composition and extinction ratios are determined, and the distribution width is varied until all three dependences on median particle radius converge. From the so obtained distribution median radius and spread, extinction values are determined for a particle number density of $N = 1$ cm^{-1} from which the true number density can then be determined by comparison with the extinction measurements.

  Hervig *et al.* (1998) assess retrieval uncertainty based on simulated extinction data and by using a Monte Carlo approach to obtain statistical standard deviations in the retrieved distributions. They determined the effective radius from the HALOE 2.45 μm extinction to within approximately $\pm$ 15% using empirical relationships

derived from *in situ* data. Uncertainties in the inferred surface areas and volumes are less than 30% and 15%, respectively.

- *Regularized Inversion Method* (RIM):
  This approach is presented by Bingen *et al.* (2002) and used to generate a global climatology of stratospheric aerosol size distribution parameters derived from SAGE II data [Bingen and Vanhellemont, 2004a; Bingen and Vanhellemont, 2004b]. In this approach the particle number density, median radius and distribution width are retrieved by minimizing a merit function, which describes the difference between previously calculated theoretical look up table values of aerosol extinction and the experimental extinction profiles normalized by the 1.02 μm aerosol extinction. The difference between both profiles is weighted by the estimated spectral experimental error and the sum is taken over the different SAGE II wavelengths. A smoothness condition is imposed on change with altitude in both median radius and distribution width, and both parameters are constrained within defined working ranges.

  Bingen *et al.* (2004b) report error estimates in median radius of approximately 50% at southern mid-latitudes in the 12-22 km altitude region and about 35% in the lower tropical stratosphere (below 20 km), in periods of reduced aerosol load. Associated uncertainties in distribution width are as high as 100 to 250%, and between about 50 and 200% in number density.

- *Linear Minimizing Error* (LME) method:
  This approach is used by Yue (1999) and Lu *et al.* (2003) to retrieve aerosol size and integrated properties, respectively. For the LME method, a random number generator is used to produce unimodal lognormal size distributions for a range of mean particle radii (0.06-0.6 μm) and distribution widths. The aerosol surface area and volume densities as well as aerosol extinction for these size distributions are calculated and stored. Aerosol surface area and volume densities are assumed to be a linear expression of the aerosol extinction available for retrieval and the coefficients in the linear expression are obtained by minimizing the retrieval errors.

- the *Anomalous diffraction approximation* (ADA):
  ADA applies to particles much larger in size than the wavelength of incident radiation. This is true for cirrus particles and for aerosols forming in the polar stratosphere. ADA has been applied by Viera and Box (1985), Franssens (1999) and Fu *et al.* (1999).

Retrieval using PCA

The operational technique used in the SAGE II data files is based on the method described by Thomason *et al.* (1997) and Steele *et al.* (1999). In the PCA approach the discretized solution of Equation 2.14 for a set of extinction measurements is given by

$$\frac{dV}{dr} = C^*(r)U\Lambda^{-1}U^*\beta, \tag{2.23}$$

where dV/dr is the vector of elements of the volume density function evaluated at radii r_j, C^* is the transpose of the volume kernel function matrix, U is the matrix containing the eigenvectors of the (kernel) covariance matrix in its columns, and Λ is the diagonal matrix of eigenvalues. The expansion is truncated by retaining only the largest eigenvalues in Λ and their associated eigenvectors in U. Integral properties of the size distribution can be written as

$$\sum_j w_j \frac{dV(r_j)}{dr}, \tag{2.24}$$

where $w_j = \Delta r$ for aerosol volume density, and $w_j = (3/r_j)\Delta r_j$ for surface area density. After substituting for dV/dr from Equation 2.23 these integral properties can be evaluated from a linear combination of the extinction measurements

$$\sum_i a_i \beta(\lambda_i), \tag{2.25}$$

where a_i are the elements of the vector given by $wC^*U\Lambda^{-1}U^*$. The coefficients a_i depend on particle composition (through the refractive index employed in the calculation of C), on the integration limits employed in the calculation of the eigenvectors and eigenvalues of the covariance matrix, and on the number of principal components retained. Steele *et al.* (1999) assessed the connection between the number of principal components retained, the choice of integration limits, and the resulting random and systematic (bias) errors. They report that, although in the absence of any experimental error the retention of all principal components results in accurate retrievals, in the presence of measurement uncertainty the coefficients are observed to propagate experimental error more readily. Reducing the number of principal components was observed to decrease error propagation, which means that a trade-off exists between the systematic error and the random error. Steele *et al.* (1999) also observed that narrowing the integration limits helps to decrease error propagation, although this means introducing another bias error. They achieved the best results when disregarding particles beyond 0.1-1.0 μm and by retaining only the two largest principal components, even though the achieved surface areas were at the same time always underestimated.

Operational SAGE II products are estimates of surface area density and effective radius. The calculation of R_{eff} was first based on the method of Thomason *et al.* (1997), using principal component analysis to derive surface area density and total aerosol volume

density from a linear combination of the four aerosol extinction measurements (where the coefficients are selected to move the weight of the retrieval to the more reliable long wavelength channels). This relationship has been simplified in the operational software using an empirical fit based on the 0.525 to 1.020 μm extinction ratio and the absolute 1.020 μm aerosol extinction that captures approximately 90% of the variance of the original [SPARC, 2006]. Only the two largest components are retained in the calculations, and the integration limits are 0.01 μm and 1.0 μm (L.W. Thomason, personal communication). A comparison between aerosol optical properties as retrieved through Optimal Estimation from the same SAGE II extinction measurements that were evaluated by NASA LaRC (using PCA) will be performed in Chapter 5 (Sec. 5.3).

The New Method

The Optimal Estimation algorithm developed here is based on Bayesian statistics. It takes a comprehensive approach to the multisolution retrieval problem by identifying all possible solutions within a given measurement uncertainty and characterizing these by means of probability density functions. An asset of this method is that from the probability density functions not only the 'best' retrieval solution (according to some criterion of optimization) can be derived but also an associated uncertainty estimate. A comparison between the OE retrieval results and those results achieved through different retrieval methods, as reported by other researchers (as described above) will be done in Section 5.1.

Chapter 3

Developing an Optimal Estimation Retrieval Model

The first step in building the new aerosol retrieval algorithm consists in expressing in a mathematical form the physical relationship between spectral measurements of light attenuation by aerosols and the aerosol properties we wish to retrieve, namely the size distribution parameters. In the second step the model is optimized to efficiently produce accurate retrieval solutions. This chapter is therefore divided into two sections. In Section 3.1 the model equations are established in a more general form by introducing the general building blocks of the algorithm. These are Bayes' Theorem, a formulation of the Maximum A Posteriori (MAP) retrieval solution, and a numerical method used to solve the matrix equations. In Section 3.2 the forward model equation is specified and the optimization procedures described. With the help of *in situ* balloon borne measurements prior knowledge can be established about stratospheric aerosol size distribution parameters, based on which a retrieval constraint can be generated. Then, the measurement and state vectors are defined, the forward model specified, and a description given of the integration procedure, the model initialisation, the minimisation procedure and ways to judge convergence.

3.1 The Bayesian Approach

In general, the relationship between a set of m measurements, arranged in a measurement vector y, and the n quantities, arranged in a state vector x, can be expressed by a vector valued forward function, f, of all influencing variables:

$$y = f(x, b) + \tilde{\epsilon}, \tag{3.1}$$

where b is a vector of retrieval parameters that influence the measurements but that will not be retrieved, and $\tilde{\epsilon}$ accounts for measurement noise. By definition, the forward

function is exact. In practice, however, the involved processes are rather complex so that $f(x, b)$ is approximated by a *forward model* $F(x, b)$. If the forward model parameters are approximately constant, Equation 3.1 becomes

$$y = F(x) + \epsilon, \tag{3.2}$$

where $\epsilon = \tilde{\epsilon} + \Delta f$ is called the *measurement error* and accounts for measurement noise as well as for the *forward model error* Δf relative to the (correct) forward function. The forward model maps a point in state space to a point in measurement space.
The retrieval solution, $\hat{x}$, is the result of operating on the measurement with some inverse method R that maps a point in measurement space to a point in state space:

$$\hat{x} = R(y, \hat{b}, x_a, \dot{c}), \tag{3.3}$$

where $\hat{b}$ is the best estimate of the forward function parameters, x_a is the *a priori* mean state which contains prior knowledge about the state vector while being unrelated to the actual measurement, and $\dot{c}$ accounts for retrieval method parameters that do not affect the measurements such as Lagrange multipliers or convergence criteria. The circumflex on the retrieved state indicates that this solution is an estimated quantity which is not necessarily identical with the true solution.
The general inverse problem can be regarded as a question of setting up and solving a set of simultaneous linear or nonlinear equations, in the presence of experimental error and often in the presence of approximations in the formulation of the equations. Each measurement comes with experimental uncertainty, resulting from limitations in the precision of the measuring instrument, and/or from a loss of information in the direct problem due to the underlying physics. (For instance, a particle with a rough surface will produce a smooth scattering pattern, and it will therefore be impossible to reconstruct the surface structure of the particle based on the scattering pattern.) Most retrieval applications have in common that the measurement error ϵ is known only statistically. Consequently, a point in state space will map into a region in measurement space, that can be described by the probability density function (pdf) of ϵ, even though the forward model $F(x)$ represents a deterministic mapping. Conversely, a measurement could be the result of a mapping from anywhere in a region of state space described by some probability density function, rather than from a single point.

Bayes' Theorem

Bayes' theory provides a formalism which, based on probability density functions, allows us to translate uncertainty in measurement space into uncertainty in state space. Bayes' Theorem relates a set of measurements, y, to the *a priori* knowledge about the required state and thereby allows us to determine the pdf of a retrieved solution state. It states

that the desired posterior pdf can be obtained by updating the prior pdf of x with the conditional pdf of a measurement:

$$P(x|y) = \frac{P(x)}{P(y)} P(y|x), \tag{3.4}$$

where

- $P(x|y)$ is the posterior conditional pdf of x which describes the probability that the state lies in the interval $(x, x + dx)$ when y has a given value,
- $P(x)$ is the prior pdf of the state x expressing quantitatively our knowledge of x before a measurement is taken,
- $P(y|x)$ is the conditional pdf of a measurement y, which describes the probability that the measurement vector lies in the interval $(y, y+dy)$ given a certain state x. To write it down explicitly requires only the statistical description of the measurement error and a forward model which expresses the measurement process by mapping the state into measurement space.
- $P(y)$ describes the knowledge about the measurement before it is taken and is in practice only a normalising factor which is often not needed [Rodgers, 2000].

This means that, given a measurement together with a description of its error statistics, a forward model describing the relation between the measurement and the unknown state, and any prior information about the unknown state, the Bayesian approach identifies all possible states that are consistent with the measured information, and assigns a probability density to them. The Bayesian approach can therefore be considered the most general approach to inverse problems [Rodgers, 2000].

The Maximum A Posteriori Solution

The Bayesian solution to the retrieval problem is the probability density function of all possible states given a particular set of measurements. As for most purposes it is desirable to select one of all possible states as 'the solution' and to assign it some error estimate, two straightforward choices are the most likely state or the expected value. The most likely state is the one for which $P(x|y)$ is maximum, whereas the expected value is the mean state averaged over the entire solution pdf:

$$\hat{x} = \int x P(x|y) dx. \tag{3.5}$$

For a Gaussian pdf, which is usually a good approximation for the experimental error and convenient for algebraic manipulation, the two options are identical because of the symmetry. This solution to the retrieval problem is called the Maximum *A Posteriori*

Solution (MAP). The (half) width of the $P(x|y)$ is a measure of the spread of all likely solutions and is therefore a measure of uncertainty of the most likely state. This means that the half width of the solution probability density function is equivalent to the one sigma uncertainty estimate of the MAP solution.

Since we are interested in retrieving several parameters from a set of measurements the transformations take place in vector space. The pdf of a vector z containing n elements with corresponding expected values $\bar{z}$ and which has a covariance matrix S_z, can be written in the form:

$$P(z) = \frac{1}{(2\pi)^{n/2}|S_z|^{1/2}} \exp\{-\frac{1}{2}(z-\bar{z})^T S_z^{-1}(z-\bar{z})\}, \tag{3.6}$$

where $P(z)\mathrm{d}z$ gives the probability that z lies within an interval limited by z and $z+\mathrm{d}z$. Surfaces of constant probability of this pdf fulfill the condition:

$$(z-\bar{z})^T S_z^{-1}(z-\bar{z}) = \text{constant} \tag{3.7}$$

and take the shape of ellipsoids in measurement space. The principal axes of these ellipsoids correspond to the eigenvectors of the covariance matrix S_z. The lengths of these principal axes are proportional to the square root of the corresponding eigenvalues and can be thought of as the multivariate equivalent of error bars. If the forward model is a general function of the state and the measurement error is Gaussian, the conditional pdf of a measurement y given some state x can be expressed in the form:

$$-2lnP(y|x) = [y-F(x)]^T S_\epsilon^{-1}[y-F(x)] + c_1 \tag{3.8}$$

where c_1 is a constant and S_ϵ is the measurement error covariance matrix. Similarly, an expression for the prior knowledge of the state can be derived, so that the pdf of the state is given by

$$-2lnP(x) = [x-x_a]^T S_a^{-1}[x-x_a] + c_2, \tag{3.9}$$

where x_a is the *a priori* mean state and S_a is the matrix containing the *a priori* state covariances. Substituting into Equation 3.4 and putting $P(y)$ into the constant leads to the general form of the Bayesian solution for the general inverse problem:

$$-2lnP(x|y) = [y-F(x)]^T S_\epsilon^{-1}[y-F(x)] + [x-x_a]^T S_a^{-1}[x-x_a] + c_3. \tag{3.10}$$

This is a quadratic form in x, which implies that it must be possible to express it as a function of a new state, namely the retrieval solution $\hat{x}$, and an associated covariance $\hat{S}$:

$$-2\ln P(x|y) = [x-\hat{x}]^T \hat{S}^{-1}[x-\hat{x}]. \tag{3.11}$$

An explicit expression for the retrival error covariance can be derived when we assume that within a small particle size range the forward model can be approximated by a linearized

forward model of the form $F(x) = \nabla_x F(x_0)(x - x_0) = y_0 + K_0(x - x_0)$, where x_0 is an arbitrary linearisation point. Equating terms that are quadratic in x then leads to an expression for the inverse covariance matrix

$$\hat{S}^{-1} = K^T S_\epsilon^{-1} K + S_a^{-1}, \tag{3.12}$$

where K is the Jacobian or weighting function matrix, that is the derivative of $F(x)$ with respect to the state vector elements. It can be observed that the (inverse) covariance matrix is the sum of contributions from both the measurement covariance and the *a priori* covariance.

If the retrieval problem was linear, an explicit expression for the solution state vector $\hat{x}$ could be derived through equating terms that are linear in x^T. As the problem is nonlinear, however, the solution state can best be found numerically, that is iteratively. In the MAP approach the expected retrieval state is situated where the posterior pdf takes a maximum, and maximising $P(x|y)$ is equivalent to finding the minimum of the righthand side of Equation 3.10, which is the same as finding the root of the multi-dimensional *retrieval cost* function

$$J = [x - x_a]^T S_a^{-1} [x - x_a] + [y - F(x)]^T S_\epsilon^{-1} [y - F(x)], \tag{3.13}$$

which at a particular state $\hat{x}$ provides a scalar valued quantity that is a measure of the goodness of the fit to both the measurements (second term on the left, or the *fit cost*) and the *a priori* state (first term on the left, or the *a priori cost*). In this form it can also be seen that the MAP retrieval solution is a weighted mean of the state and the *a priori* state. However, if the *a priori* pdf is appropriate, the solutions will be biased only within experimental uncertainty. Searching for the root leads to the following implicit expression for $\hat{x}$

$$-\hat{K}^T S_\epsilon^{-1} [y - \hat{F}] + S_a^{-1} [\hat{x} - x_a] = 0, \tag{3.14}$$

where $\hat{K}$ is the Jacobian matrix of derivatives at the solution state. The challenge of the retrieval model is to locate one of the multiple minima of the multidimensional cost function which is as close as possible to the global minimum. The method employed here is the Levenberg-Marquardt method, a root finding method specifically designed for non-linear least squares problems. This method combines two approaches:

- The *Gauss-Newton* approach, which is an approximation of the Newton method for moderately nonlinear problems. It works best near the solution where the cost function tends to be more quadratic. For a cost function which is exactly quadratic in x the minimum will be found in one step.
- The *steepest descent* method, which calculates the local gradient of the function in order to step downhill at each iteration; this approach still works reasonably far from the solution, where it therefore replaces the *Gauss-Newton* approach.

If $g(x)$ denotes the implicit solution expression for x (Eq. 3.14) the general form of the Levenberg-Marquardt iteration is:

$$x_{i+1} = x_i - [\nabla_x g(x_i) + \gamma I]^{-1} g(x_i) \tag{3.15}$$

where I is the identity matrix and γ a Lagrangian multiplier that controls the weight of the steepest descent method. As the state vector elements have different magnitudes and dimensions, it is reasonable to normalize by S_a^{-1} instead of by I. Substituting Equation 3.14 for $g(x)$ and dropping the second derivative of the forward model leads to the equation:

$$x_{i+1} = x_i + (S_a^{-1} + K_i^T S_\epsilon^{-1} K_i + \gamma S_a^{-1})^{-1} (K_i^T S_\epsilon^{-1} [y - F(x_i)] - S_a^{-1} [x_i - x_a]). \tag{3.16}$$

The Lagrangian multiplier, γ, is chosen at each step to minimise the cost and is updated such that the new value of x remains within the linear range of the previous estimate, the so called *trust region*. This depends very much on the nature of the problem and is therefore a matter for experiment.

3.2 Refining the Model

In order to enable the new retrieval model to perform successfully the algorithm has to be tailored to the characteristics of the aerosol retrieval problem. This includes deciding on the measurement and state vectors, developing an efficient and accurate mathematical formulation of the forward model, formulating an *a priori* constraint, implementing an effective minimization procedure, identifying a suitable model initialization, and developing effective convergence criteria. The overall aim of the new algorithm is to retrieve aerosol microphysical properties accurately while keeping the computational effort within reasonable limits.

Prior Information from Measured Data

Useful information about the size ranges and size distributions of stratospheric aerosol particles can be gained from prior measurements. The *a priori* information about size distribution parameters used in this research originates from *in situ* balloon borne size resolved aerosol concentration measurements collected above Laramie, Wyoming (41°N, 105°W), using the University of Wyoming aerosol particle counters. From the measured cumulative concentrations, functional size distributions were derived (by the University of Wyoming team) by fitting the measured data with either monomodal or bimodal lognormal size distributions. The fitting method consists of minimizing the root mean square error $rmse = \sum_i \log^2[N_m(> r_i)/N(> r_i)]$, where $N_m(> r_i)$ is the measured concentration of all particles with radii greater than a particular radius r_i, and $N(> r_i)$ associated number density resulting from integrating the lognormal size distribution over all particles

with $r > r_i$. All combinations of measured concentrations are tried to find the set of 5 discrete sizes plus condensation nuclei (CN) giving size distribution parameters which minimize the root mean square error when all measurements are included. If there are less than five independent concentration measurements then a unimodal distribution is applied, otherwise bimodal distributions are used. The fitted size distributions are, however, not forced to be bimodal. If a unimodal distribution provides a better fit, then that is used instead of a bimodal fit, but this is rarely the case when there are more than four independent measurements [Deshler *et al., 2003*]. The Wyoming aerosol data are available on the anonymous ftp site: trex.uwyo.edu [Deshler *et al., 2003*].

The 30 vertical profiles of size distribution parameters used here originate from data measured between 1 May 1991 and 22 October 1997 and cover altitudes between 20 km and 35 km. Figure 3.1 presents the lognormal distribution width as a function of the particle median radius for monomodal and bimodal aerosol size distributions. It can be observed that particles of monomodal size distributions and particles forming the first mode of the bimodal distributions are very similar. Sizes typically range between a few nanometres and 0.4 μm (in radius), with the majority of particles measuring between 0.02 μm and 0.2 μm. The associated distribution widths range between approximately 0.1 and 1 (in units of log-radius). Particles in the second mode are generally larger and typically measure between 0.2 and 1.0 μm. The associated distribution widths are smaller than in the case of the first mode, and are less than 0.4 for the majority of the data. In general the distribution width is observed to decrease with increasing particle size. This can be explained by the growth processes of the aerosol particle populations. Basically, if mass is approximately conserved or not enough material is available to let all particles grow equally, a few larger ones will grow at the expense of many small particles.
Figure 3.2 presents the same data as Figure 3.1, but separated by mode and in the form of histograms (of the logarithmic values), and supplemented by the associated number densities. Additional Gaussian distributions are shown, which were generated based on the mean and standard deviation of the data (and normalized to highest value of the associated histogram). These Gaussians lognormal distributions will be used in the retrieval algorithm to describe prior knowledge about the size distribution parameters. It can be observed that in the case of monomodal aerosols (frames a, d, g) both the Gaussians and the histograms are symmetric and that the widths of the lognormal distributions exceed those of the histograms. The number densities of bimodal aerosols (frames b and c) tend to be less symmetric than the monomodal distributions. This is likely to be due to the growth dynamics. The time scales on which changes occur are relatively short in volcanically perturbed atmospheres (bimodal distributions), whereas background distributions tend to reflect an equilibrium state.

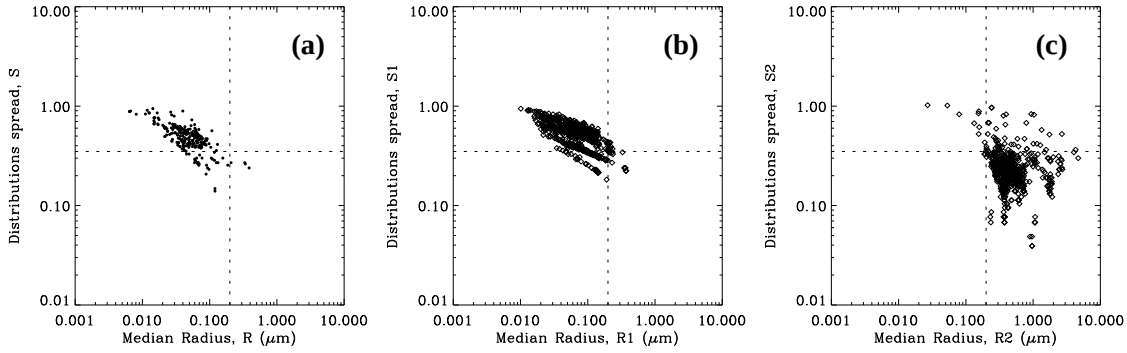

Figure 3.1: Median radius R versus lognormal size distribution width S of (a) monomodal, (b) first mode of bimodal, and (c) second mode of bimodal distributions, as measured by the University of Wyoming optical particle counters over Laramie (41°N) between 1991 and 1997. The dotted lines at 0.2 μm in radius and 0.35 in spread were drawn as grid lines to facilitate the cross comparison between the different modes.

Measurement Vector:
The measurement vector, y, used in our forward model consists of a set of four volume extinction coefficients – one for each of the four SAGE II aerosol spectral channels. Although, strictly speaking, β^{ext} is not a measurement as such but derived from the raw data, it will nevertheless be referred to as 'measurement' in this document.

State Vector:
Since we are interested in the aerosol microphysical properties, the state vector should contain the particle size distribution parameters in some form. In the case of monomodal (that is background) aerosol the total number density, median particle radius and distribution width are approximately normally distributed (see Fig. 3.2.a/d/g). Consequently, a suitable state vector with respect to the Bayesian approach would be

$$x = \ln[N, R, S], \tag{3.17}$$

where ln denotes the natural logarithm. The logarithmic form also has two more advantages. Firstly, the largely different size ranges and orders of magnitude of N, R, and S as observed above (Fig.s 3.1 and 3.2), namely approximately 1-100 cm^{-3} particles, 0.001-1.0 μm in radius, and 0.1-1 in log radius (monomodal) are brought to a more similar scale by using the logarithms. Secondly, the solution space of a logarithmic state vector naturally constrains the retrieval to physical solutions, as negative numbers in log-space translate to positive values of number density, median particle radius and size distribution width.

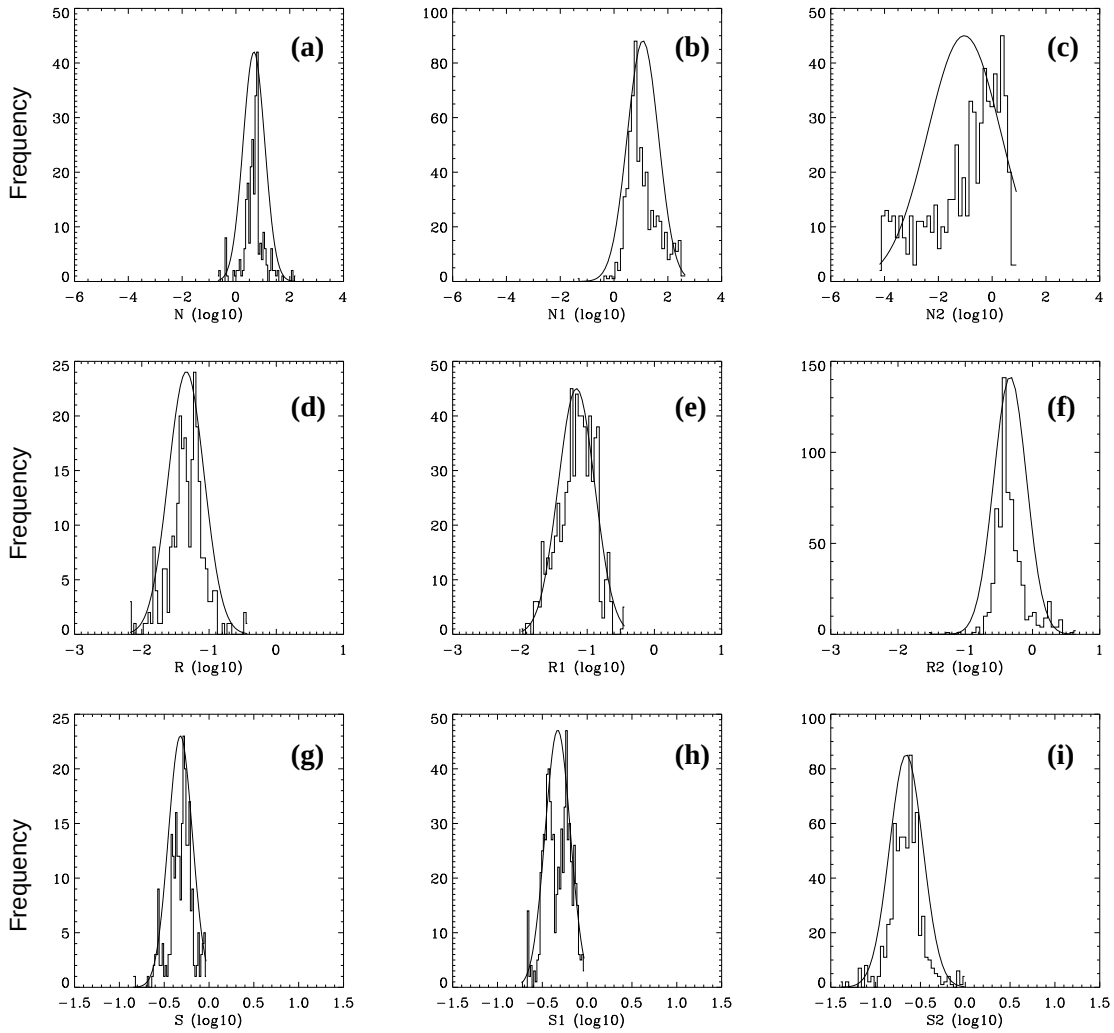

Figure 3.2: Histograms of monomodal (a, d, g) and bimodal size distribution parameters as generated from *in situ* measurements of background aerosol taken over Laramie by the University of Wyoming between 1991 and 1997. The associated Gaussian distributions were generated based on the mean and standard deviation of the measurements, and scaled to match the highest frequency of the associated histogram.

A Priori Constraint:

In the Bayesian approach, the retrieval solutions can be constrained by expressing our prior knowledge about the state vector elements in the form of a probability density function $P(x)$. Using all *in situ* measurements of monomodal size distribution parameters, as shown above, an *a priori* mean state

$$x_a = \ln\left[\bar{N}, \bar{R}, \bar{S}\right] \;=\; \ln[4.7, 0.046, 0.48] \;=\; [1.54, -3.08, -0.73] \tag{3.18}$$

and an associated covariance matrix

$$S_a = \begin{pmatrix} 0.86 & 0.06 & 0.03 \\ 0.06 & 0.38 & -0.14 \\ 0.03 & -0.14 & 0.10 \end{pmatrix}$$

can be generated. The diagonal elements (S_{ii}) and the off-diagonal elements (S_{ij}) were calculated according to

$$S_{ii} = \sigma_{\mathrm{i}}^2 \; = \; \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3.19}$$

$$S_{ij} = \sigma_{\mathrm{ij}} \; = \; \frac{1}{n-1} \sum_{i=1,j=1}^{n} (x_i - \bar{x})(x_j - \bar{x}) \tag{3.20}$$

$$\tag{3.21}$$

where n=3 is the number of state vector elements, i= [1,2,3] denotes the association with $[N, R, S]$, σ_{i}^2 are the variances, and σ_{ij} are the covariances. The prior pdfs $P(x) = P(x_a)$ can be generated based on the mean *a priori* state vector elements and the associated standard deviations $\sigma_a = [\sigma_{\ln N}, \sigma_{\ln R}, \sigma_{\ln S}] \; = \; [0.93, 0.61, 0.31]$. They are identical with the lognormal functions shown in Figure 3.2.a/d/e. As these pdfs are wider than the associated histograms they can be considered a conservative description of the measured distributions. By using these pdfs as *a priori* constraint we aim to achieve physically sensible values without constraining the solution space too tightly.
Since $d \ln z = dz/z$, the *a priori* uncertainties (in %) of the linear variables $[N, R, S]$ are the same as the absolute uncertainties of the logarithmic variables, that is

$$[\sigma_N, \sigma_R, \sigma_S] = [93, 61, 31]\%. \tag{3.22}$$

An Efficient and Accurate Forward Model: The extinction of light by aerosol particles can be be described by Equation 2.13, which expresses the extinction coefficient as a continuous integral of extinction cross-sections over all particle sizes between zero and infinity. In practice, the solution space can be reduced to some finite integration limits $r_a > 0$ and $r_b < \infty$, and the integral to a sum of discrete terms or function evaluations at a given number of grid points. The forward model (in log radius space) is then

$$F(x) = \beta^{ext}(\lambda) \; = \; \sum_{\ln r_a}^{\ln r_b} \pi r^2 \cdot Q^{\mathrm{ext}}(r, \lambda, RI) \cdot \frac{dN(r)}{d \ln r} \Delta \ln r, \tag{3.23}$$

where r is a vector containing the particle sizes at which the function is evaluated, RI denotes the refractive index, and $\Delta \ln r$ is the width of the particle size interval in natural log space. Making the algorithm efficient means reducing the number of function evaluations

of the mathematically complex and hence computationally expensive Mie calculations and merging the upper and lower integration limits r_a and r_b without jeopardizing the accuracy of the integral.
In theory, the largest possible particle size is determined by the balance between buoyancy and gravitational settling, so $r_{b,max}$ is of the order of a few micrometers [Seinfeld and Pandis, 1998]. The smallest aerosol particles are condensation nuclei with radii of the order of a few nanometres. In practice, the integration interval can be reduced even further by tailoring $[r_a, r_b]$ individually to each particle size distribution ($x = \ln N, R, S$). As the shape of the integrand in Equation 3.23 is largely determined by the lognormal particle size distribution function, and as calculating the extinction coefficient is mathematically complex, it was decided to estimate suitable integration limits based on the integrand divided by Q^{ext}. Both $\ln r_a$ and $\ln r_b$ are determined analytically as those particle sizes where the function reduces to 1% of its maximum value.
Within these limits the integral is initially evaluated at 17 discrete grid points that, in log space, are evenly distributed between r_a and r_b. New grid points are then added to the middle of each interval and the integral (Eq. 3.23) is repeatedly evaluated, until the total of differences (at all four spectral wavelengths) between two successive integral evaluations is smaller than 0.1% of the latest value: $|\beta_n - \beta_{n-1}|/\beta_n < 0.1\%$. This means that the *discretization error* can expected to be much smaller than 1%. Since for the size range of monomodal aerosol particles the integrand of Equation 3.23 is reasonably smooth and the function is mathematically well behaved, a classical adaptive integration procedure can be applied which is based on the extended or composite trapezoidal rule and on Simpson's rule [Press *et al., 1992*]. At every point in radius space the function is evaluated only once and stored, so it can be reused in all following iterations. The integration starts with $N_0 = 17$ points, and there are usually less than 5 iterations per retrieval process. Suitable integration limits can be determined as follows: Those radii, where the function $\tilde{f} = \pi r^2 * n(r)$ reduces to a approximately 1% of its maximum, are taken as r_a and r_b.

Retrieval Initialisation: The algorithm retrieves aerosol properties from one set of four spectral extinction values at a time which describes an aerosol at one particular location (and time). In order to generate vertical profiles of aerosol properties aerosol extinction data are evaluated one height at a time. Here, the data are stored such that they are analysed one profile after the other and always starting at the lowermost height.
As the aerosol retrieval problem is strongly nonlinear and has multiple solutions, a good retrieval initialisation may save retrieval time as well as increase the chance to converge to the correct solution. One option is to use available prior information about the solution space and to chose the *a priori* mean state as a first guess, that is the starting point in state space. Where successive measurements form vertical profiles, the retrieval solution at one height can be used as a first guess for the retrieval analysing the data measured

at an adjacent height. Here, the *a priori* mean state is used as first guess in the very first retrieval process. In all following retrieval runs the solution of the preceding retrieval process is used as a starting point for the following retrieval (that is the analysis of the next set of measurements). This seems a reasonable approach, since adjacent measurements within a vertical profile are likely to be reasonably close to each other. In case no solution is found, the following retrieval process (analysis of the next data set) will use the *a priori* mean state as a starting point. As the current scheme does not (yet) distinguish the measurements belonging to different vertical profiles, it happens that the first retrieval within a new profile (bottom height) is initialized with the solution state of the uppermost height of a different profile. This is not generally a problem, but it will be seen later (in Sec. 4.5) that here is probably some potential for further improvement. Since, however, all in all the current scheme works well, any further refinements of the initialization will be left to future studies.

Efficient Minimisation: The efficiency of the Levenberg-Marquardt root finder scheme (in minimizing the cost function) is determined by the way the Lagrangian multiplier γ (see Eq. 3.16) is updated. This procedure depends on the nature of the retrieval problem and is therefore a matter for experiment. In order to make a step toward a new state, γ has to be changed after each iteration according to whether the state has been improved or not. This is judged based on the change of the cost function (Eq. 3.13), where generally, a decrease in J reflects improvement. Marquardt (1963) suggested updating γ through multiplying or dividing by 10 at each iteration. Here, multiplication by 10 and division by 20 (of $\tilde{\gamma} = (1 + \gamma)$) was found to be more suitable.
Ideally, the new state lies within the linear range of the model. There is, however, no guarantee that this is actually the case, and therefore one or several elements of the new state may fall beyond physically sensible values. In order to prevent this, (which would slow the model down tremendously or prevent it from converging altogether), a penalty function is implemented in the inverse model which checks if particle densities are between 0.01 and 1000 cm^{-3}, median radii between 0.001 and 5 μm, and lognormal distribution widths between 0.01 and 1.5. If a state vector element falls below the lower threshold it is reset to the minimum value; if it exceeds the upper threshold it is brought back to the maximum allowed value, then the retrieval process continues from there.

Defining Convergence: Judging convergence means to decide whether a retrieval process has converged to sufficient precision yet or whether further iterations are necessary. It is not the same as assessing the accuracy (correctness) of the achieved results, which is the content of a later chapter (Sec. 4.3). In general, convergence criteria can either be developed based on the absolute size of a variable or based on the change of quantities

between two subsequent iterations. In our model, combinations of the following criteria are used:

1. **Cost**: The retrieval cost (Eq. 3.13) is an important retrieval diagnostic as it is the variable to be minimised. Although good retrieval solutions are generally accompanied by a small cost, the absolute size of the cost does not identify whether a retrieval has converged or not. The change in the cost function between subsequent iterations, however, will always be small near a cost minimum. In addition, it appears sensible to ask that, at the final solution, the difference $\Delta J = J_i - J_{i+1}$ be positive, (which means that the cost is decreasing), and an order of magnitude smaller than the expected value of the cost function (which in the case of SAGE II is $m = 4$). A suitable convergence criterion is therefore $0 < \Delta J/m < 0.25$ ($0 < \Delta J < 1$), where the threshold 0.25 (1.0) was found experimentally to work well.

2. **Measurement vector**: The difference in the retrieved extinction vectors between two subsequent iterations should be smaller than the difference $\delta\hat{y}$ between the retrieved signal, $F(\hat{x}) = \hat{y}$, and the measurement, y. The associated covariance $S_{\delta\hat{y}}$ can be shown to be $S_{\delta\hat{y}} = S_\epsilon(\hat{K}S_a\hat{K}^T + S_\epsilon)^{-1}S_\epsilon$ [Rodgers, 2000]. Since the difference between two subsequent iterations $|F(x_{i+1}) - F(x_i)| = |y_{i+1} - y_i| = \Delta y$ has an expected value of $m = 4$, a good convergence criterion is $(\Delta y^T S_{\delta\hat{y}}\Delta y)/m < 0.1$ or $\Delta y^T S_{\delta\hat{y}}\Delta y < 0.4$.

3. **Size distribution parameters**: The change in the linear variables (as opposed to the logarithmic state vector elements) can be compared to the associated retrieved uncertainties. The difference in number density, median radius, and distribution width has to be smaller than the retrieved uncertainties.

4. **Iterations**: Generally, a maximum of 60 iterations is allowed, because models that are used to process large amounts of data have to be fast and efficient. In practice, this upper limit is hardly ever reached. However, it helps to end processes that do not converge within a reasonable amount of time.

Which of those criteria are applied depends on the number of iterations, i, performed. At more than 10 (but less than 60) iterations, the retrieved solution is accepted if criterion 1 or criterion 2 is fulfilled. At less than 10 iterations, criteria 1 and 3, or criterion 2 have to be met. When 60 iterations are reached and the convergence criteria are still not met, the process is stopped and the result rejected. The (combination of the) above convergence criteria were tested and found to work fast and efficiently. Solutions are found fast and mostly in less than 5 (and always less than 60) iterations. This means that the algorithm is fast enough to process large amounts of data.

Chapter 4

Model Validation with Synthetic Data

In this chapter, the performance of the new retrieval algorithm is assessed based on simulated spectral aerosol extinction. For this purpose, retrievals are performed on two sets of synthetic extinction data that differ only by their respective noise levels:

- (a) the *Minimum Noise Scenario* (minNS) is characterised by a 1% Gaussian distributed random noise component added onto each of the four spectral extinction data;
- (b) the *Maximum Noise Scenario* (maxNS) is characterised by [60, 45, 30, 25]% Gaussian distributed random noise on the respective spectral channels [0.385, 0.452, 0.525, 1.020] μm.

These particular values form the lower and upper ends of a range of experimental SAGE II extinction uncertainties (see Fig. 2.6). The first scenario (minNS) is meant to show how well the algorithm works in the presence of very little extinction uncertainty, whereas the second scenario (maxNS) assesses how well the the OE retrieval model performs in the presence of very large experimental uncertainty. With a realistic measurement uncertainty distribution (similar to that shown in Fig. 2.6) the quality of the retrieved aerosol properties is expected to be somewhere between the two validation scenarios (see Ch. 5, Sec. 5.1).

This validation chapter includes the following aspects: Section 4.1 starts by introducing the testbed for the retrieval validation, which consists of synthetic extinction data at the four SAGE II spectral channels. The information content (with respect to aerosol size distribution parameters) of these 'synthetic measurements' is then determined, which demonstrates the necessity of retrieval constraints. In Section 4.2 all retrieved results (and the associated correct solutions) are used to develop a quality filter, which allows the

identification of unsuccessful results and to separate those from the more likely retrieval solutions. In Section 4.3 all retrieved aerosol properties and the associated retrieved uncertainties are presented and described. Section 4.4 contains a comprehensive error analysis. It includes the validation of the retrieved uncertainties by comparing them to the true errors, an introduction to possible additional retrieval error sources and the quantification of all important error components. In Section 4.5, a small number of vertical profiles as formed from the retrieved aerosol properties are compared with the associated correct profiles. In this form we gain indications as to why some retrieved results are better than others.

4.1 The Testbed

The 264 aerosol size distributions used here to generate synthetic aerosol extinction originate from the same *in situ* measurements that were described earlier (see Sec. 3.2). These data were measured at midlatitudes (41°N) between approximately 15 and 40 km and form vertical profiles of number density, median particle radius and distribution width. Based on these data extinction coefficients were calculated for a 75% (by weight) sulphuric acid solution at 300 K and four spectral wavelengths (0.385 μm, 0.452 μm, 0.525 μm, and 1.020 μm). The specifications of the two test data sets are summarized in Table 4.1.

Wavelength (μm)	0.385, 0.452, 0.525, 1.02
Wavenumber (cm^{-1})	26001, 22104, 19048, 9807
Refractive Index	[Palmer and Williams, 1975]
Composition	H_2SO_4-H_2O
Concentration (% by weight)	75
Temperature (K)	300
Distribution Type	monomodal
1. Noise Scenario (% per λ)	1, 1, 1, 1
2. Noise Scenario (% per λ)	60, 45, 30, 25

Table 4.1: Specifications of the synthetic aerosol data, that will be used to validate the performance of the new retrieval algorithm. The noise levels give the mean values of a Gaussian distributed random noise at the associated wavelength (spectral channel).

Spectral Dependence

In order to be able to retrieve the three state vector elements from SAGE II data, each spectral signature has to be associated with a particular combination of lnR and lnS

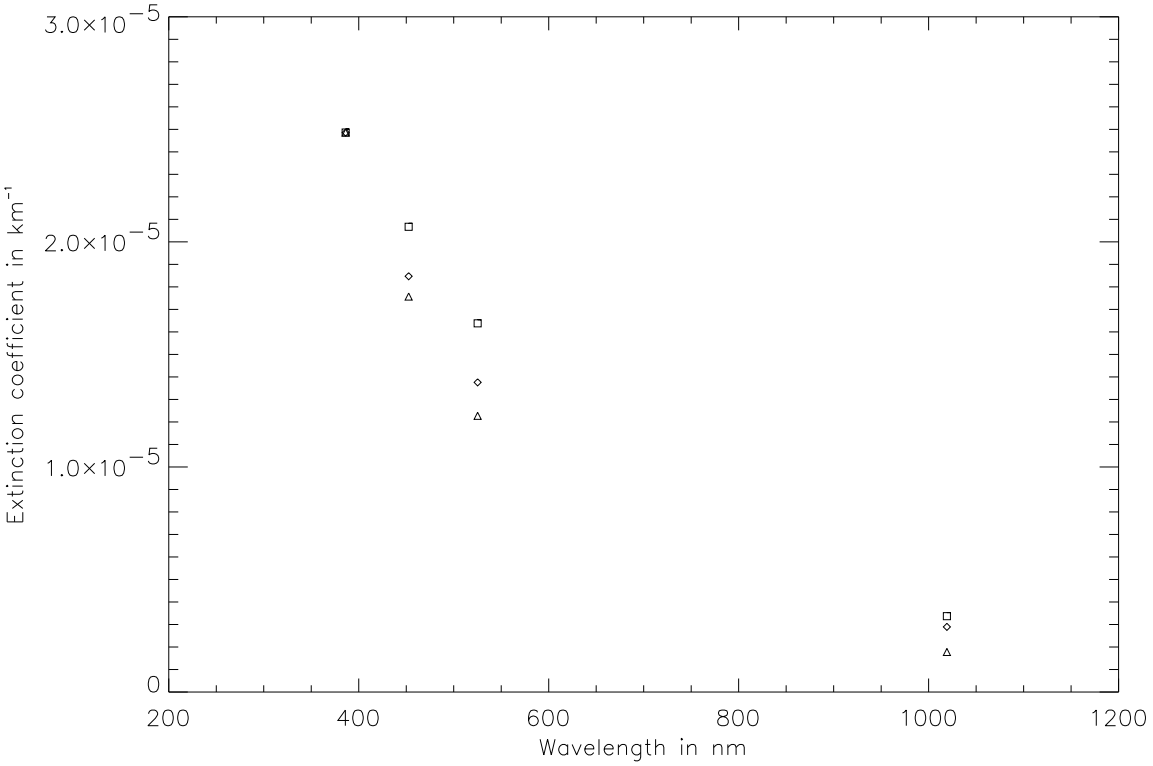

Figure 4.1: Spectral dependence of the extinction coefficient β^{ext} (in km^{-1}) of three typical combinations of median particle radius and distribution width: small ($R = 0.008$ μm; $S = 0.90$, diamonds), medium ($R = 0.067$ μm; $S = 0.45$, triangles) and large ($R = 0.183$ μm; $S = 0.25$, squares) background aerosols. The three sets were linearly scaled (through changing N) to match at the shortest wavelength.

and lnN. Figure 4.1 shows the spectral signatures of three combinations of lnR and lnS describing small, medium, and large background aerosols. The three sets of spectral measurements were scaled linearly to match at the shortest wavelength. It can be observed that the spectral dependence is different for all three aerosols, which indicates that the size distribution parameters can be retrieved independently.

Figure 4.2 shows the dependence of the spectral aerosol extinction coefficients on median particle radius and lognormal distribution width. It can be observed that for the three typical combinations of R and S (marked by plus signs here) there is a clear dependence on both variables, which confirms that R and S of typical background aerosols can be retrieved independently. Toward smaller widths the dependence of β^{ext}on S decreases. Comparison with Figure 3.1, however, indicates that such combinations of R and S have not been measured.

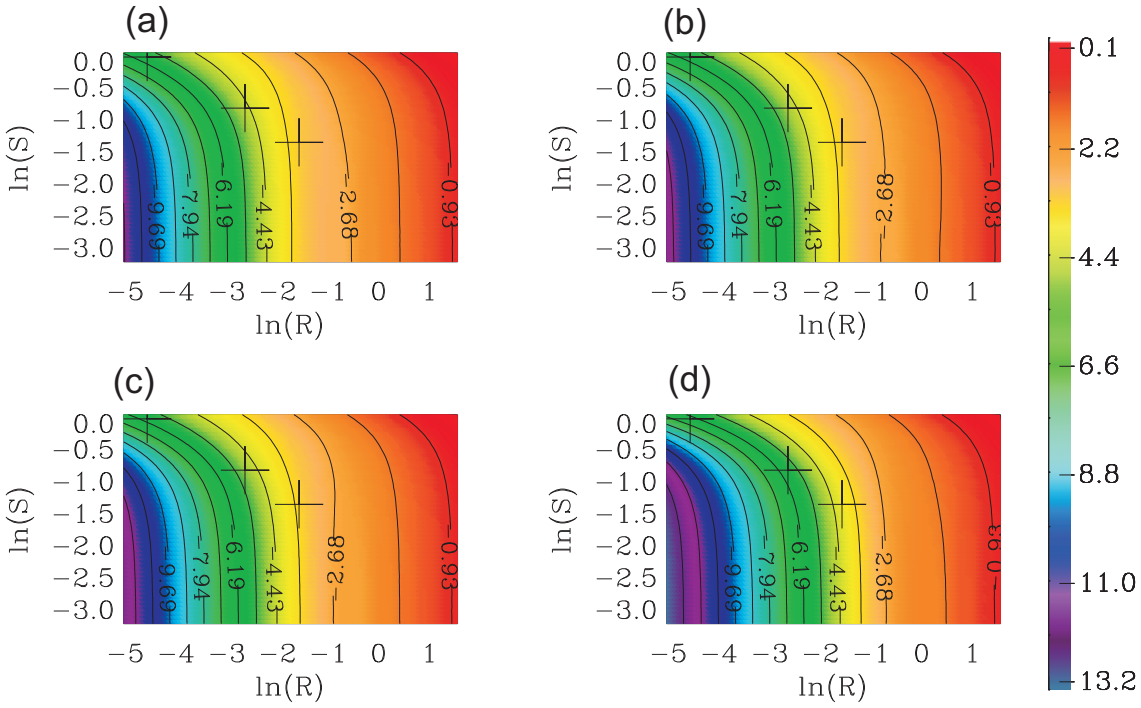

Figure 4.2: Extinction (in log10) as a function of median particle radius ($R\epsilon[0.005, 5]$ μm) and lognormal distribution width ($S\epsilon[0.05, 1]$ in units of log radius) at (a) 0.385 μm, (b) 0.452 μm, (c) 0.525 μm, and (d) 1.020 μm. Three example distributions are indicated by '+': small ($R = 0.008$ μm; $S = 0.90$), medium ($R = 0.067$ μm; $S = 0.45$), and large ($R = 0.183$ μm; $S = 0.25$) background stratospheric aerosols. (The total number density is $N = 1$ in all cases).

Information Content of the Measurement Vector

Useful retrieval results can only be achieved if the measurements contain information about the quantities we wish to retrieve. It is therefore useful to quantify the information content of the spectral extinction measurements that form the measurement vector. Information content can be quantified in various ways. Two widely used measures of information are the *Shannon Information Content* and the *Degrees of Freedom* of a measurement. Both are measures of the number of independent pieces of information that can be retrieved from the measurements in the presence of measurement uncertainty.

The Shannon definition expresses information content in a general qualitative sense as the factor by which knowledge of a quantity x can be improved by measuring y. Because it is a scalar quantity it is very useful in optimising observing systems as well as characterising and comparing them [Rodgers, 2000]. The Shannon information content, H, arises from information theory and describes the change in entropy S of a system between probability

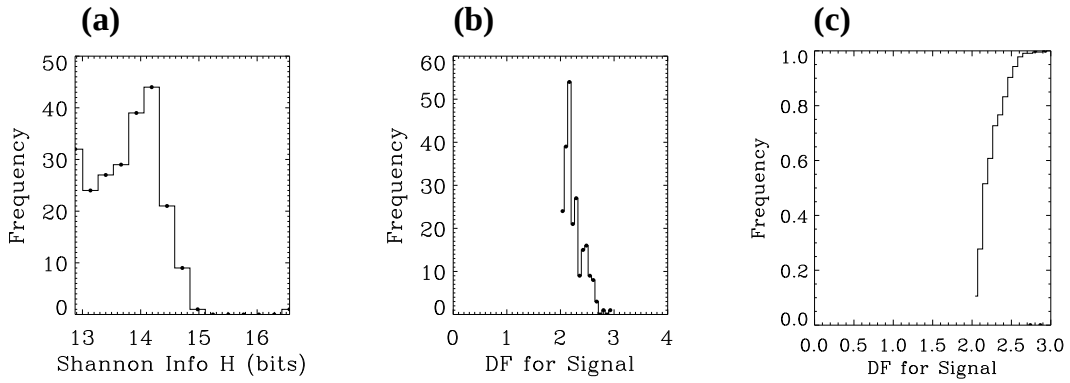

Figure 4.3: Minimum Noise Scenario: Histograms of the (a) Shannon Information Content, H, and (b) the degrees of freedom for signal, d_s. Frame (c) presents the d_s in the form of a cumulative frequency distribution.

density functions before ($P(x)$) and after taking a measurement ($P(x|y)$). In analogy to thermodynamics, where the entropy of a system is defined as the logarithm of the number of internal states of the system that give rise to the same measured macro-state, the Shannon information content of aerosol extinction data is a measure of the number of different states x that with a given noise level can be distinguished by the measurement system. The concept is a generalisation of the signal-to-noise ratio. Rodgers (2000) shows the entropy of a system is related to its covariance matrix and derives the following expression for H based on Gaussian statistics:

$$H = \frac{1}{2}\ln|S_a| - \frac{1}{2}\ln|\hat{S}| = -\frac{1}{2}\ln|\hat{S}S_a^{-1}| \tag{4.1}$$

where S_a and $\hat{S}$ are the prior and posterior covariance matrices of the aerosol state vector, and $|...|$ indicates the determinant. Conventionally, H is given in units of bits, which means that the natural logarithm in the above equation is replaced by log to the base 2. H will always be positive and $\hat{S}$ smaller than S_a provided that the new measurement is drawn from the underlying pdf.

Figures 4.3.a and 4.4.a show frequency distributions of H for both test data sets, i.e. with small (minNS) and large noise levels (maxNS). In all cases H is positive which shows that the retrieval covariance is always smaller than the prior covariance, meaning that the prior knowledge was improved through the measurement. Also, as expected, the noisier data provide less information. With large noise H is typically about 4 bits compared to 13 to 15 bits in the minNS, which corresponds to $2^4 = 16$ and $2^{13} = 8192$ or $2^{15} = 32,768$ distinguishable states. This illustrates clearly that with increasing experimental noise the retrievable information is reduced as the ability to identify the correct state decreases.

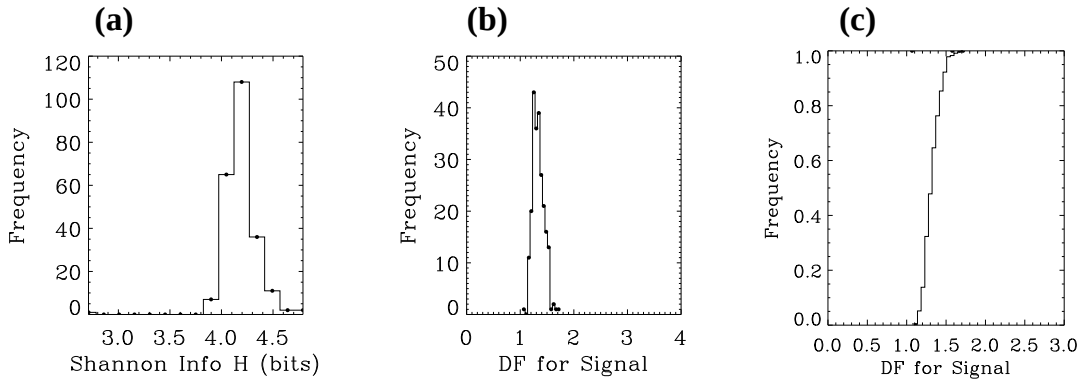

Figure 4.4: As Fig. 4.3 but for the Maximum Noise Scenario.

The number of degrees of freedom of the signal, d_s, is of more direct use for this study as it expresses how many independent quantities are retrievable in the presence of experimental noise. An expression for the degrees of freedom can be derived from the Jacobian K, and it can be shown (see Rodgers, 2000) that d_s can be calculated from the singular values, λ_i, of the transformed Jacobian $\tilde{K} = S_\epsilon^{-\frac{1}{2}} K S_a^{\frac{1}{2}}$:

$$d_s = \sum_{i=1}^{m} \frac{\lambda_i^2}{1+\lambda_i^2} \tag{4.2}$$

Figures 4.3.b/c and 4.4.b/c present histograms and cumulative distributions of the degrees of freedom for signal. It can be observed that with little noise (Fig. 4.3.b) there are generally between 2 and 3 independent quantities, in 90% of the cases there are less than 2.5 (Fig. 4.3.c). With large measurement uncertainty (Fig. 4.4.b) d_s drops down to between 1 to 2, in 90% of all cases less than 1.5 independent pieces of information (Fig. 4.4.c) are retrievable. Since for monomodal lognormal size distributions three retrievable quantities are required, there is an obvious lack of information which will have to be compensated for by the use of prior information in the form of *a priori* constraints.

Noise scenario	minNS	maxNS
Retrieval Results	absolute / relative in %	absolute / relative in %
Total Analysed	264 / 100	264 / 100
Converged	259 / 98	264 / 100
Accepted Solutions	231 / 88	232 / 88
Retrieved vs True	correl. coeff.	correl. coeff.
$\ln N$	0.56	0.52
$\ln R$	0.86	0.80
$\ln S$	0.85	0.70
$\ln A$	0.98	0.94
$\ln V$	1.00	0.98
$\ln R_{\text{eff}}$	0.93	0.90

Table 4.2: Model validation: Number of analysed, converged and accepted retrieval solutions (in absolute/relative numbers), and correlation coefficients describing the linear correlation between all accepted ('good') and the associated correct aerosol properties. Given the large number of measurements these correlation coefficients are all significant.

4.2 Quality Filter

From a total of 264 retrievals, all (maxNS) or nearly all (minNS) processes converged to a solution. In order to be able to automatically distinguish the good and bad solutions, a quality filter of some sort is required. With knowledge of the correct solutions it was found that the total cost and the averaging kernel are suitable diagnostics to form a filter. The aim was to achieve a good balance between maximizing the correlation coefficient

$$r = \frac{\sigma_{\text{xz}}}{\sigma_{\text{x}}\sigma_{\text{z}}} = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (z_i - \bar{z})^2}} \tag{4.3}$$

between the remaining retrieved results x_i and the associated true solutions z_i, (where σ_{xz} denotes the covariance between the x_i and the z_i, and σ_x and σ_z are the standard deviations) and minimizing the data loss through rejection. Table 4.2 lists the resulting number of accepted solutions and associated correlation coefficients after applying the following constraints, which (used in combination) were found to successfully identify measurements with a comparatively small information content and retrieval solutions that differ significantly from the true state:

1. all diagonal averaging kernel elements must be smaller than 2.

2. the retrieval cost has to be smaller than 20.

It can be observed that after applying this quality filter, 88% of all retrievals are accepted as 'good' results, regardless of the noise level. All correlation coefficients are highly significant in the sense that the probability that the same number of results of two uncorrelated variables give higher correlations is less than 0.05% (see Taylor, 1939, Table C, Appendix C). The correlation coefficients are observed to be considerably lower in number density than in median particle size and distribution width, and they are highest in surface area and volume densities. These numbers indicate that it is harder to correctly retrieve the lower order moments than the higher order moments of the size distribution. This is likely to be a consequence of the particle number density and size being more sensitive to the smallest particles (to which aerosol extinction at visible wavelengths is practically invisible) than surface area density or volume density (see Sec. 1.2).

In order to double-check the new filter from a statistical point of view, a χ^2-test was performed based on the difference $\delta y = (y - \hat{y})$ between all accepted solutions $\hat{y} = F(\hat{x})$ and the associated synthetic extinction y. A general description of the χ^2 significance test is given in Appendix A. Appendix B describes its application to δy. Although the δy are not strictly Gaussian distributed (not shown), the χ^2 test gives an indication as to whether or not the differences are significant and hence whether or not the quality filter is able to identify the bad retrieval solutions.
On performing the test we find approximately 20% (minNS) and 7% (maxNS) of all χ^2 values to be significant at the 5% level. This means that with (unrealistically) little noise the filter should ideally have rejected more solutions, and with large noise (maxNS) the achieved result is close to the ideal 5%. (Since the χ^2 of $\hat{y}$ increases with decreasing covariance matrix $S_{(\hat{y}-y)}$ (see Eq. B.3), and as $S_{(\hat{y}-y)}$ decreases with decreasing extinction uncertainty (see Eq. B.2), it is indeed likely that in the minimum noise scenario more differences δy will be found to be significant than in the maximum noise scenario.) We conclude that the χ^2-test was able to confirm that the new *ad hoc* quality filter performs well.

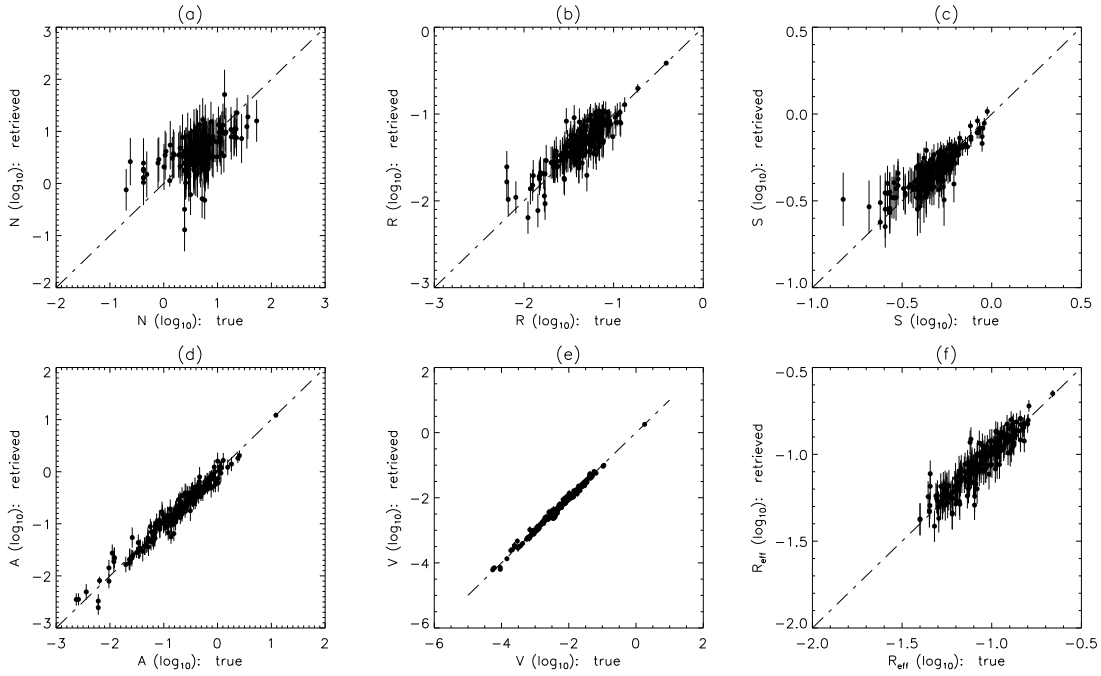

Figure 4.5: Minimum Noise Scenario: True versus retrieved particle number density (a), median radius (b), distribution width (c), and associated surface area density (d), volume density (e), and effective radius (f), with their respective uncertainties. The broken line indicates the one to one line.

4.3 Retrieved Aerosol Properties

Figures 4.5 (minNS) and 4.6 (maxNS) show the retrieved aerosol properties of all accepted solutions as a function of the associated correct values. It can be observed that the integrated aerosol properties are closer to the correct solutions than the retrieved size distribution parameters from which they were derived. It can also be observed that under large noise conditions the smallest values of all retrieved size distribution parameters tend to overestimate the correct values (Fig. 4.6.a-d). In the Minimum Noise Scenario (Fig. 4.5) such a trend is not noticeable. This indicates that under large noise conditions (maxNS) the smallest values are likely to be biased by the larger influence of the *a priori* constraint. The size of the error bars is observed to be a good description of the general spread of all solutions about the diagonal broken line, which means that retrieved and the correct solutions mostly agree within the error bars. This applies to both noise scenarios, even though the retrieved uncertainties are generally larger with large noise (maxNS, Fig. 4.6) than with very little noise (minNS, Fig. 4.5). The retrieved uncertainties will be looked at more closely in Section 4.4.

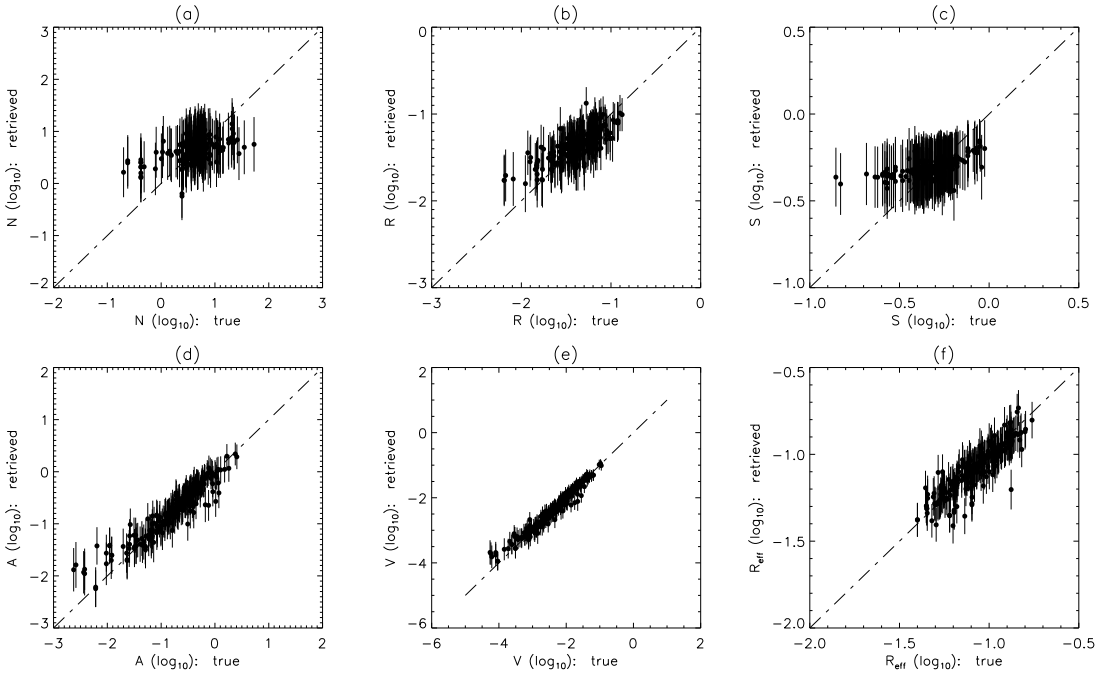

Figure 4.6: As Fig. 4.5 but for the Maximum Noise Scenario.

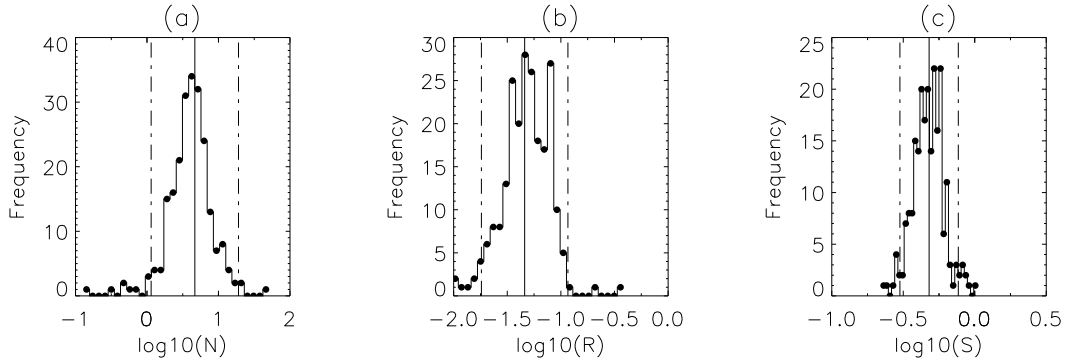

Figure 4.7: Minimum Noise Scenario: Histograms of the retrieved number density (a), median radius (b) and distribution width (c). The vertial lines mark the *a priori* state (solid) and plus/minus the one sigma *a priori* standard deviations (dash-dot).

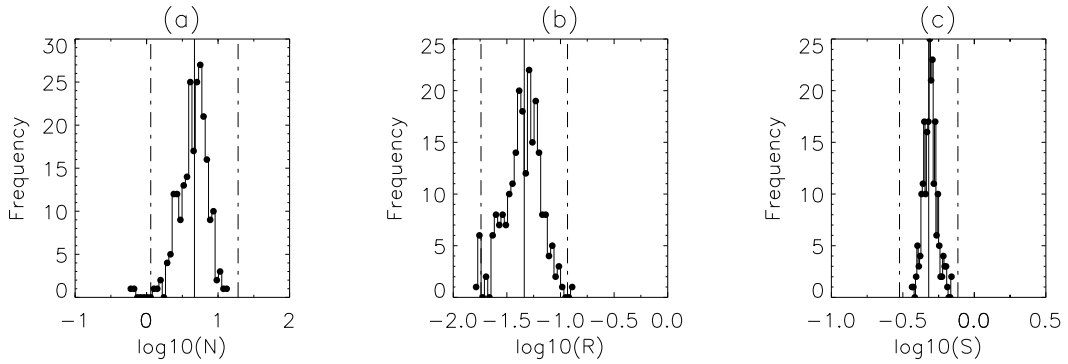

Figure 4.8: As Figure 4.7 but for the Maximum Noise Scenario.

In Figures 4.7 (minNS) and 4.8 (maxNS) the retrieved solutions are shown in the form of histograms. Comparison with Figure 3.2.a/d/g shows that the retrieved solutions are very close to the correct size distribution parameters. Generally, all histograms are approximately symmetrical about the *a priori* value and the great majority of all solutions fall within the *a priori* mean state plus or minus one standard deviation. The observation that with large noise (maxNS) the solution ranges are somewhat narrower than in the Minimum Noise Scenario is most likely a consequence of the stronger influence of the *a priori* constraint. The symmetry about the *a priori* mean, as found in both the correct and retrieved solutions, was very much expected and clearly desired here, because the correct solutions are the same values based on which the *a priori* constraint was generated. Deviations from this symmetry can be expected when analysing data that are entirely independent of the *a priori* values, as, for instance, in the case of SAGE II measurements of aerosol extinction (see Ch. 5, Sec. 5.1).

Figure 4.9 and Figure 4.10 present histograms of all retrieved uncertainties. With very little experimental noise (Fig. 4.9), the retrieved errors are typically between 40 and 75% in number density, between 15 and 30% in median radius, and between 5 and 25% in distribution width. With large experimental noise (Fig. 4.10, maxNS), the uncertainties are mostly between 70 and 85% in N, between 30 and 50% in R, and between 20 and 30% in S, such that the lower bounds happen to match roughly with the upper bounds in the minimum noise scenario.
In Figures 4.11 and 4.12 the uncertainties of the derived integrated aerosol properties are presented. With minimum noise (minNS), typical uncertainties are observed to be 10-30% in surface area density, 5-20% in volume density, and 5-15% in effective radius. With large extinction uncertainty (maxNS), the uncertainties are naturally higher and typically 35-50% in A, 25-50% in V, and 15-19% in R_{eff}.
Table 4.3 lists the associated ensemble mean retrieved (relative) uncertainties and thus

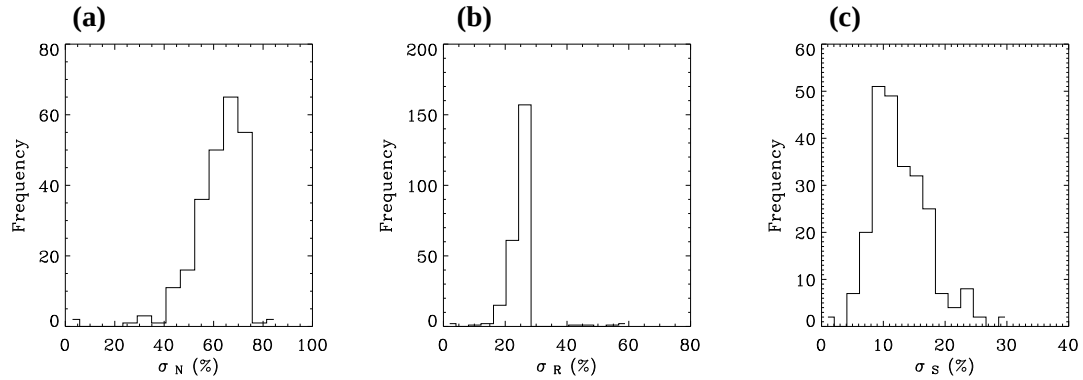

Figure 4.9: Minimum Noise Scenario: Retrieved (one sigma) uncertainties in (a) number density, (b) median radius, and (c) distribution width.

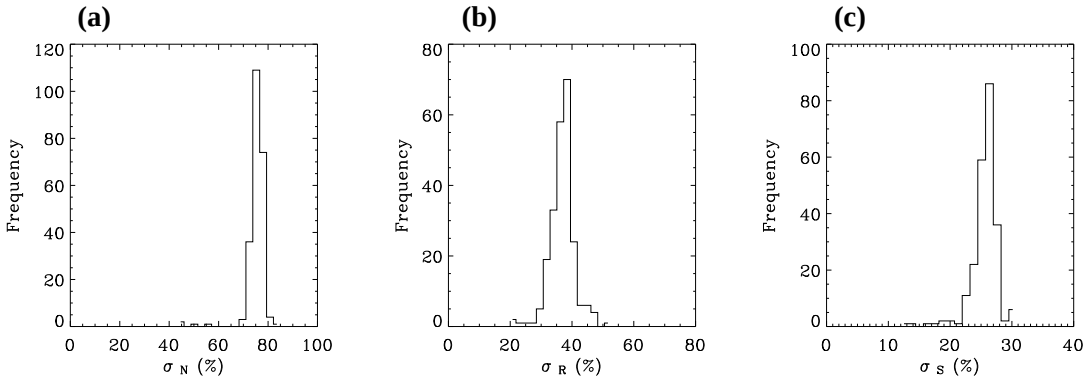

Figure 4.10: Same as Figure 4.9 but for the Maximum Noise Scenario.

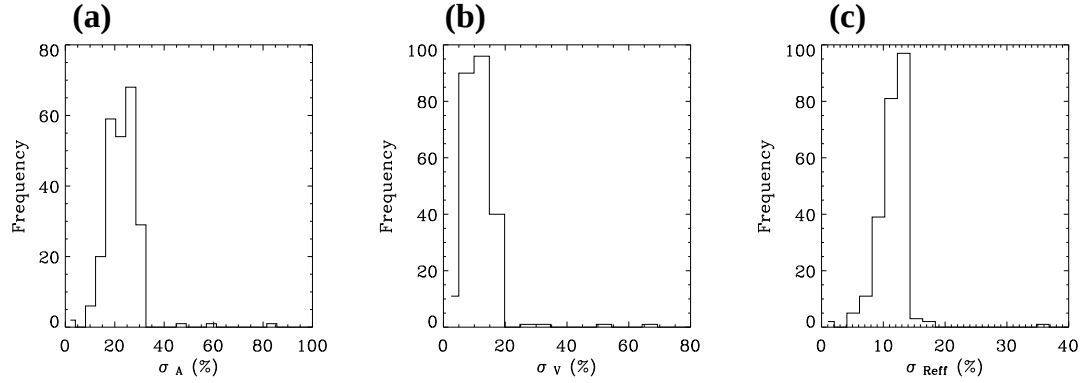

Figure 4.11: Minimum Noise Scenario: Relative uncertainties (in %) in surface area density (a), volume density (b), and effective radius (c) as derived from the retrieved size distribution parameters.

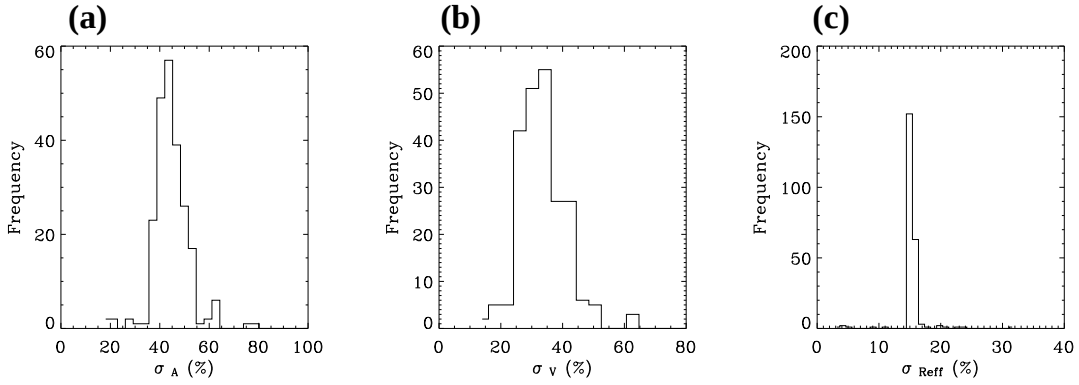

Figure 4.12: Same as Fig. 4.11 but for the Maximum Noise Scenario

	little noise (minNS)	large noise (maxNS)
	σ_N, σ_R, σ_S (%)	σ_N, σ_R, σ_S (%)
Ensemble Mean	62, 24, 14	75, 37, 26
	σ_A, σ_V, σ_{Reff} (%)	σ_A, σ_V, σ_{Reff} (%)
Ensemble Mean	22, 11, 11	45, 34, 15

Table 4.3: Ensemble mean retrieved relative uncertainties (in %) in number density, median radius, distribution width, surface area density, volume density, and effective radius.

provides an overview of the typical orders of magnitudes under the different noise conditions. A detailed analysis of the retrieved uncertainties, the various components, and of other possible additional error sources, is presented below (Sec. 4.4).

4.4 Error Analysis

The objective of the following error analysis is to understand the various sources of retrieval uncertainty and how they propagate into the final product. Section 4.4.1 starts by looking at the sensitivity of the retrieval model, which determines how the various uncertainties propagate into the retrieval solution. Important diagnostics in this respect are the model bias, the gain and the averaging kernel. Section 4.4.2 quantifies the contributions to the retrieved uncertainties from measurement noise and from the *a priori* constraint and estimates the forward model error and the forward model parameter errors. In Section 4.4.3 sdditional sources of uncertainty are described. This includes estimating the effect and order of magnitude of bimodal errors that can arise under aerosol background conditions. Section 4.4.4 shows a direct comparison between the retrieved uncertainties and the true errors (that is the difference between the retrieved and the correct aerosol

properties). The orders of magnitude of the different errors are summarized and compared in Section 4.4.5.

4.4.1 Model Sensitivity

The sensitivity of the model can be described by how the solution is related to the true state and, through the retrieval constraint, to the *a priori* state. In terms of the forward model linearised about the *a priori* state, the conceptual retrieval solution given in (Eq. 3.3) can be written as

$$\hat{x} = R[F(x_a, \hat{b}) + K_x(x - x_a) + K_b(b - \hat{b}) + \epsilon, \hat{b}, x_a, c], \tag{4.4}$$

where R represents the transfer function, $K_x = \partial F/\partial x$ and $K_b = \partial F/\partial b$ are the partial derivatives of the forward model to the state and to the forward model parameters, respectively, x is the true solution state, and b the true forward model parameters. Linearisation of the transfer function with respect to its first argument leads to the expression

$$\hat{x} = R[F(x_a, \hat{b}), \hat{b}, x_a, c] + \frac{\partial \hat{x}}{\partial y}[K_x(x - x_a) + \epsilon]. \tag{4.5}$$

where the error is now defined as $\epsilon = K_b(b - \hat{b}) + \Delta f + \tilde{\epsilon}$ combining contributions from forward model parameter errors (fmp error), from the forward model error (fm error) and from measurement noise.

$$G = \frac{\partial \hat{x}}{\partial y} = \frac{\partial R}{\partial y} \tag{4.6}$$

is the *retrieval gain* matrix which, mathematically, is a generalised inverse of the weighting function matrix K. The retrieval gain describes the sensitivity of the retrieved state to deviations in the measurements. In the one-dimensional case it is also known as the *contribution function*. Since the product $G\epsilon$ is the measurement error translated into solution space, the size of the gain gives an indication of the ill-conditioning of a solution method. The larger G the more the errors will be amplified during the retrieval process. From Equation 4.5 it follows that the sensitivity of the retrieval solution to the true state x is given by a matrix product

$$A_v = \frac{\partial \hat{x}}{\partial x} = GK_x, \tag{4.7}$$

which is called the *averaging kernel* matrix A_v. The averaging kernel also provides information about the influence of the *a priori* state on the solution, as will be detailed later. In terms of the difference between the retrieved and the *a priori* states Equation 4.5 can be expanded to express the difference between the retrieved and the true states (the latter being identical to the *a priori* state here)

$$\hat{x} - x_a = (R[F(x_a, \hat{b}), \hat{b}, x_a, c] - x_a) + A_v(x - x_a) + G\epsilon, \tag{4.8}$$

where the three terms on the right are the *retrieval bias*, a smoothing term, and the contribution from total error ϵ, respectively. The retrieval bias or *retrieval method error* is the residual of a simulated error-free retrieval of the *a priori* state and is the error in the solution due to imperfections in the retrieval method. The retrieval bias is zero for a well behaved inverse method, which means that (at least to first order) the retrieval method is not sensitive to any retrieval method parameter other than x_a. Without the bias term Equation 4.8 can be rearranged to

$$\hat{x} = xA_v + x_a(I - A_v) + G\epsilon, \tag{4.9}$$

which expresses that the Maximum A posteriori Probability Optimal Estimation solution, $\hat{x}$, is essentially a weighted mean of the true (yet unknown) and the *a priori* states plus a (small) contribution from measurement error. From this it directly follows that the sensitivity of the retrieval method and solution to the *a priori* state is

$$\frac{\partial R}{\partial x_a} = \frac{\partial \hat{x}}{\partial x_a} = I - A_v. \tag{4.10}$$

Consequently, if the averaging kernel is equal to the identity matrix, then the *a priori* state has no influence on the solution, which instead is purely retrieved from the measurements. In the case of the SAGE II measurements, the averaging kernels are expected to deviate from the identity matrix due to the limited information content.

Results

The retrieval bias (see Eq. 4.8) or *retrieval method error* can be examined by calculating the residual of a simulated error-free retrieval of the *a priori* state. Therefore, the error-free spectral extinction (at SAGE II wavelengths) of the *a priori* aerosol size distribution (see Sec. 3.2) was generated. The retrieval was then performed for a range of initial conditions including very large and very small first guess states in order to ensure a robust test. We found that all retrievals converged toward the correct solution, x_a, and that the bias was always zero within rounding error, namely $< 10^{-6}$. It can therefore be concluded that, at least to first order, the sensitivity of the retrieval method to any inverse model parameters (other than the *a priori* state) is negligible.

The retrieval gain or error amplification factor is found to be relatively large (of the order of 10^4), which is due to the non-linearity of the problem. This means that a small difference in measurement space can bring about a significant change in the retrieved state (aerosol properties). The large gain is most likely a consequence of the linear dependence between the spectral SAGE II measurements and the ill-posed nature of the aerosol retrieval problem [Steele *et al.*, 1999]. A large gain does not mean the retrieval problem cannot be solved, but it underlines the difficulty of the retrieval problem at hand.

The ensemble mean and standard deviations of the diagonal averaging kernel elements (of the validation retrievals) are listed in Table 4.4. As expected (as a consequence of the

	$A_{v,ii}$ (minNS)	$A_{v,ii}$ (maxNS)
Mean	0.52, 0.77, 0.97	0.24, 0.77, 0.30
Standard deviation	0.16, 0.15, 0.19	0.05, 0.20, 0.21

Table 4.4: Model validation: Ensemble mean and standard deviation of the diagonal averaging kernel elements for both the minimum (minNS) and the maximum Noise Scenarios (maxNS). The $A_{v,ii}$ are associated with number density, median radius, and distribution width, respectively.

limited information content), they are all smaller than 1, which means that there is clearly an influence of the *a priori* state on the retrieved solutions. This influence is largest for number density and amounts on average to between 48% (minNS) and 76% (maxNS). The influence of the *a priori* state is observed to be, on average, about 23% on particle median radius, and between 3% (minNS) and 70% (maxNS) on the distribution width. As expected, the averaging kernel elements decrease with increasing noise levels as the influence of the *a priori* constraint becomes more important.

4.4.2 Retrieval Error Components

Equation 4.9 can be expanded and rearranged to give the true solution error as the difference between the retrieved ($\hat{x}$) and the correct (x) states

$$\begin{aligned} \hat{x} - x &= x_a(I_n - A_v) + x(A_v - I_n) + G\epsilon \\ &= (I_n - A_v)(x_a - x) + GK_b(b - \hat{b}) + G\Delta f + G\tilde{\epsilon}. \end{aligned} \tag{4.11}$$

The first term on the right side is the *smoothing error* which arises from the influence of the *a priori* constraint. The other three error components arise from forward model parameter errors, from forward model error, and from retrieval noise. In practice, $\tilde{\epsilon}$ is substituted for ϵ. This means that the retrieved uncertainty takes into account the measurement noise and smoothing errors only, whereas the forward model error and forward model parameter errors have to be estimated by other means.

Smoothing Error

The smoothing error depends on the information content (with respect to the state vector) of a particular set of measurements. The smaller the information content, the larger the influence of the *a priori* constraint and the associated smoothing error, as indicated by the averaging kernel matrix. With synthetic measurements, where the true solution state x is known, the smoothing error $(A_v - I)(x - x_a)$ is straightforward to evaluate with I being the $n \times n$ identity matrix. In the case of measured data the actual smoothing error cannot be estimated because the true state is unknown. Instead, the statistics of the error can be characterised by using the mean and covariance over some appropriate ensemble of

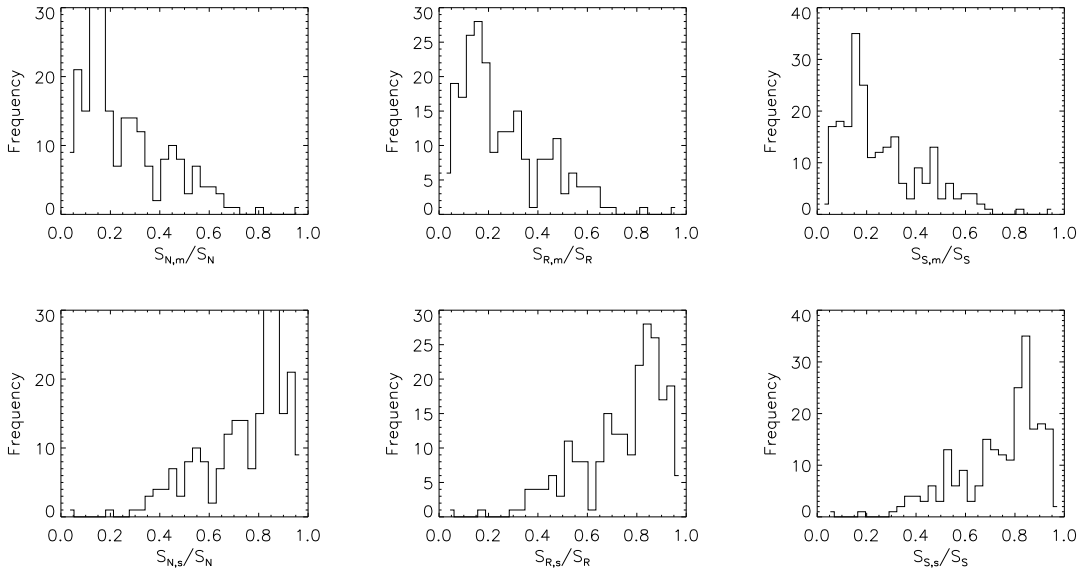

Figure 4.13: With minimum noise (minNS): Histograms of the fractional contributions to the retrieved uncertainty $\hat{S}$ from smoothing error $S_m/\hat{S}$ (top) and from measurement noise $S_s/\hat{S}$ (bottom).

states which may or may not be that described by x_a and S_a [Rodgers, 2000]. The mean is $(I - A_v)(x_a - \bar{x})$, which becomes zero if an ensemble has been chosen for which $\bar{x}$ is equal to x_a. Then, if S_e is the covariance of the ensemble of states about the mean state, the covariance of the smoothing error about $\bar{x}$ is:

$$\begin{aligned} S_\mathrm{s} &= \varepsilon\{(A_v - I)(x - \bar{x}) \cdot (x - \bar{x})^T (A_v - I)^T\} \\ &= (A_v - I) S_\mathrm{e} (A_v - I)^T. \end{aligned} \tag{4.12}$$

Retrieval Noise

The *retrieval noise* is, through the gain matrix, largely determined by the nonlinearity of the model. It is equal to $G\tilde{\epsilon}$ and is the most straightforward of the error components to evaluate, with synthetic as well as with measured data. With a known measurement noise covariance matrix S_ϵ, the retrieval noise covariance S_m (the square root of the diagonal elements of which form the standard deviation of the retrieval noise) can easily be calculated as a matrix product with the gain matrix G (Eq. 4.6):

$$S_\mathrm{m} = G S_\epsilon G^T. \tag{4.13}$$

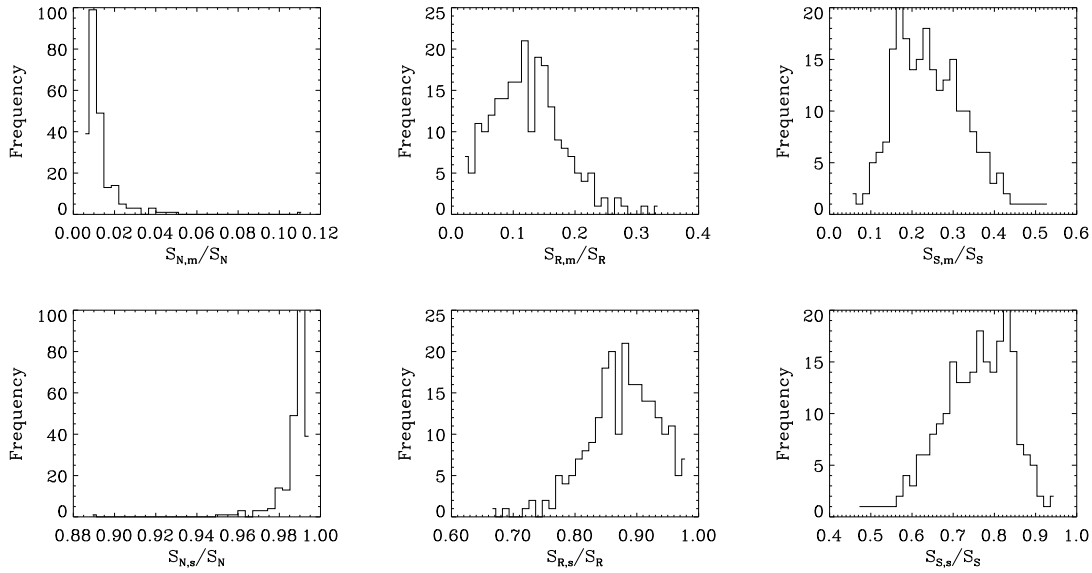

Figure 4.14: As Fig. 4.13 but for the Maximum Noise Scenario

Contributions of smoothing error and retrieval noise to the retrieved uncertainty

Figures 4.13 and 4.14 present histograms of the fractional contributions to the retrieved uncertainty $\hat{S}$ from retrieval noise, $S_m/\hat{S}$, and from smoothing error, $S_s/\hat{S}$. It can be observed that generally the smoothing error accounts for the major part of the retrieved uncertainty and that the retrieval noise component tends to be small. Comparison between the two different noise scenarios shows that the influence of the smoothing error component on $\hat{S}$ grows with increasing extinction uncertainty, as the influence of the *a priori* constraint becomes more important. Cumulative histograms (not shown) of the fractional contributions presented in Figures 4.13 and 4.14 show that with little extinction uncertainty (minNS), the contribution to $\hat{S}$ from the retrieval noise is $\leq 30\%$ and from smoothing error is $\geq 70\%$ in 68% of all retrievals. With large extinction uncertainty (maxNS), the respective contributions to $\hat{S}$ are $\leq 5\%$ (from S_{m}) and $\geq 95\%$ (from S_{s}) in 68% of all results.

Forward Model Parameter Error

The forward model parameter error or parameter random error arises from uncertainties in parameters which are not part of the state vector (to be retrieved) but nevertheless influence the measurements. A common approach is to use best-guess values, the uncertainty of which is represented by a random deviation of the 'true' atmosphere about this guess. Theoretically, propagation of forward model parameter uncertainties into the

Reference states	(a)	(b)
T (K)/c (weight %)	220/70	200/65
m(T,c) at 0.368 μm	1.45520	1.45527
m(T,c) at 0.453 μm	1.44799	1.44780
m(T,c) at 0.525 μm	1.44276	1.44240
m(T,c) at 1.020 μm	1.43077	1.43009

Table 4.5: Test scenarios for Forward Model Parameter Error estimates. The refractive index $m(T, c)$ is listed for four spectral wavelengths for both temperature/concentration scenarios.

modeled extinction and eventually into the retrieved product, can be derived analytically ($K_b = \partial F/\partial b$), or by perturbation of the uncertain parameter in the forward model. In the retrieval model at hand the main non-retrievable parameter is the aerosol refractive index $m(T, c)$ (Eq. 2.16). Uncertainties in the aerosol refractive index can arise from errors in atmospheric temperature and water vapour partial pressure, and in the associated error in aerosol concentration $c(T, p_e)$.

The forward model parameter error can be assessed by observing the impact of fluctuations in temperature and aerosol composition on the refractive index and the accuracy achieved in the retrieved aerosol properties when using those perturbed model parameters. For this purpose two common scenarios of temperature and sulphuric acid concentration were chosen, that serve as 'true' reference states: (a) 220 K/70%, and (b) 200 K/65%. The associated refractive indices at 0.386, 0.453, 0.525, and 1.020 μm are listed in Table4.5. For both temperature/concentration scenarios, noise-free spectral extinction data were generated based on a medium sized aerosol with N= 4.7 cm^{-3}, R= 0.04 μm and S=0.48 (which is the *a priori* mean state). From those extinction data, monomodal aerosol size distribution parameters were then retrieved, assuming different combinations of temperature and concentration fluctuations therein: (i) $\pm$1 K (in temperature) and $\pm$1% (in concentration), or (ii) $\pm$5 K and $\pm$5%. The largest resulting differences to the true reference state are listed in Table 4.6. It can be observed that the forward model parameter error is of the order of a few percent ($< 3\%$). This is generally more than an order of magnitude smaller than the retrieved uncertainties (Tab. 4.3), which suggests that the forward model parameter error tends to be negligible compared to the smoothing error and retrieval noise components which together form the retrieval error covariance $\hat{S}$.

Forward Model Error

The forward model error is the difference between the exact physics and the description by the (imperfect) mathematical model. For a given atmospheric state there will be a system-

Temperature (K) Concentration (%)	220±1 70±1	220±5 70±5	200±1 65±1	200±5 65±5
Δm (%)	0.01	0.10	0.60	0.50
$\Delta \ln N$ (%)	0.08	0.39	1.66	0.45
$\Delta \ln R$ (%)	0.04	0.18	0.42	0.20
$\Delta \ln S$ (%)	0.03	0.12	0.73	0.14
ΔN (%)	0.12	0.60	2.51	0.68
ΔR (%)	0.11	0.55	1.29	0.62
ΔS (%)	0.02	0.01	0.53	0.10

Table 4.6: Forward model parameter error estimates for a medium sized (N= 4.7 cm^{-3}, R= 0.04 μm, S=0.48) sulphuric acid aerosol and two temperature/concentration scenarios. The listed numbers are the maximum differences in the refractive index, Δm, and in the retrieved variables relative to the true reference state. The error estimates are given for both the logarithmic and the linear variable space.

atic difference between spectra calculated using the forward model and true spectra from the atmosphere. It can in practice only be evaluated if the true physics and instrument response of the modeled system are known in their entirety. With the model at hand, errors may arise from any discrepancies that may exist between Mie scattering theory and the true physics, for instance due to the atmospheric particles not being homogeneous and spherical, the size distribution parameters not being lognormally distributed, or due to numerical errors such as the truncation in the range of particle sizes considered or the discretization of the mathematical functions involved.
Since in the case of stratospheric background aerosol at temperatures above the frost point, the deviation of the tiny sulphuric acid droplets from the particle's sphericity is practically zero [Tolbert, 1994] and the particles can be assumed to be homogeneous, the Mie solutions should be close to exact. Assuming that the deviations of monomodal particle size distribution parameters from the lognormal function are small (Fig. 3.2.a/d/g), the forward model error of the model at hand can be approximated by the numerical errors due to discretization of the model equations and due to truncation of the integration integral (see Sec. 3.2).

Exact quantification of the truncation error is not straightforward owing to the mathematical complexity of the forward model equations. However, since near the upper and lower integration limits the forward model function is determined by the lognormal particle size distribution, the error can be estimated based on the the lognormal integral. If the integration limits are defined by where the Gaussian function equals one hundredth of its maximum, then the disregarded tails of the function amount to approximately 0.24%

of the main integral, that is of the extinction value. This is a conservative estimate, as multiplication of the Gaussian function with r^2 and Q^{ext} (Eq. 2.13) reduces the function near the left tail even faster to zero, thereby reducing the size of the truncated tail. At the large particle end the multiplication of the Gaussian size distribution with $r^2 \cdot Q^{\text{ext}}$ has a similar effect, as $r^2 \ll 1$ and $Q^{\text{ext}} \to 2$.
The error due to discretization of the forward model functions is smaller than the difference between the second last iteration and the last integral β^{ext} in a retrieval process, which by definition has to be smaller than 0.1% of the last extinction value (see Sec. 3.2). This results in a total forward model error of less than 1%. This is close to the order of magnitude of typical SAGE II extinction errors (see Fig. 2.6) at 1.020 μm (which are typically about 1-10%), but considerably smaller than the noise at 0.368 μm (which is typically 10-60%). This means that the forward model error tends to be negligible compared to the retrieved uncertainties at 0.368 μm, but not necessarily compared to those at 1.020 μm.

4.4.3 Other Sources of Uncertainty

Additional uncertainty in retrieved monomodal aerosol parameters can arise when the evaluated extinction measurements in reality originate from bimodal aerosol and not from monomodal aerosol. Although, in general, non-volcanic stratospheric aerosols are well described by monomodal particle size distributions, Deshler *et al.* (2003) suggest that a number of their aerosol counts measured *in situ* could indicate bimodal distributions, even though the second modes are generally small. Since the Optimal Estimation retrieval model at hand is set up to retrieve monomodal distribution parameters only (due to the limited information content), analysis of extinction from bimodal aerosols can lead to uncertainties, which will be referred to as *bimodal errors*.
Additional solution uncertainty arises when clouds contaminate the measurements. This happens predominantly in the cold polar stratosphere, where the ambient temperature falls below the frost point and polar stratospheric clouds form, or else in the tropical tropopause region, where tropospheric clouds rise high into the tropopause region or even penetrate into the stratosphere.
Retrieval errors due to bimodal aerosol or clouds can in theory be avoided by distinguishing between extinction caused by monomodal aerosol and extinction caused by bimodal aerosol or clouds. As will be shown below, it is very difficult or impossible to distinguish between extinction signals caused by monomodal and bimodal aerosols. Therefore, the magnitude of bimodal errors (under stratospheric background conditions) will be estimated. Extinction by monomodal and bimodal size distributions will be called 'monomodal extinction' and 'bimodal extinction', respectively.

Distinguish monomodal from bimodal extinction

The tests are performed on synthetic extinction data based on monomodal and bimodal aerosol data as measured over Laramie/Wyoming under atmospheric background conditions (see Prior Information, Sec. 3.2). Figure 4.15 displays monomodal extinction (panel a) and bimodal extinction (panel b) at two different wavelengths. Comparison of this data suggests that separating the two based on their magnitude is not straightforward here. Both the monomodal and bimodal extinction are observed to cover nearly the same size range, so that the data are practically indistinguishable in this respect.It can, however, also be observed that extinction at 0.525 μm tends to be higher than that at 1.020 μm, and that this difference is more distinct in the bimodal case. This indicates that there is a different wavelength dependence between small and large particles. Kent *et al.* (1993) have used this dependence to distinguish aerosol particles from the larger cloud droplets in SAGE I data. They observed that the ratio of extinction at 0.435 μm to that at 1.02 μm lay between 2 and 5 for pure background aerosols and below 2 in the presence of cloud particles. Following this approach, extinction ratios were formed (between 0.525 μm and 1.020 μm data) and presented in Figure 4.15.c in the form of a normalised cumulative distribution. Here, we observe that monomodal ratios, which are generally associated with smaller particles, tend to be larger than the bimodal ratios. Nevertheless, the range is identical, which means that the ratio cannot be used to distinguish between different aerosols.

Bimodal Error

In order to quantify bimodal errors, the above extinction data, which were simulated using bimodal size distributions, were used to retrieve monomodal aerosol properties. In Figure 4.16 the retrieved surface area densities, volume densities, and effective radii are presented and compared with the correct bimodal values. It can be observed that the retrieved surface area densities (Fig. 4.16.a/d) are close to the correct bimodal values, but tend to underestimate the latter at surface areas greater than approximately 1 $\mu m^2 cm^{-3}$. The retrieved monomodal volume densities (Fig. 4.16.b/e) are also observed to be very close to the true bimodal volumes with a slight tendency of the monomodal solutions to underestimate the bimodal values. In effective radius (Fig. 4.16.c/f), the monomodal results tend to overestimate the bimodal values, (which indicates that the retrieved surface areas are more strongly underestimated than the retrieved volume densities). The linear correlation coefficients between the retrieved monomodal and the correct bimodal values are 0.89 in A and 0.94 in V, with associated probabilities of getting a similar value for uncorrelated values of less than 0.05% (see Taylor, 1939, Appendix C). This means that the achieved correlation coefficients are highly significant. The low correlation coefficient in effective radius (0.29) is most likely caused by one outlier (see Fig. 4.16.c). An alternative test which assesses how many retrieved values agree with the correct values to within the

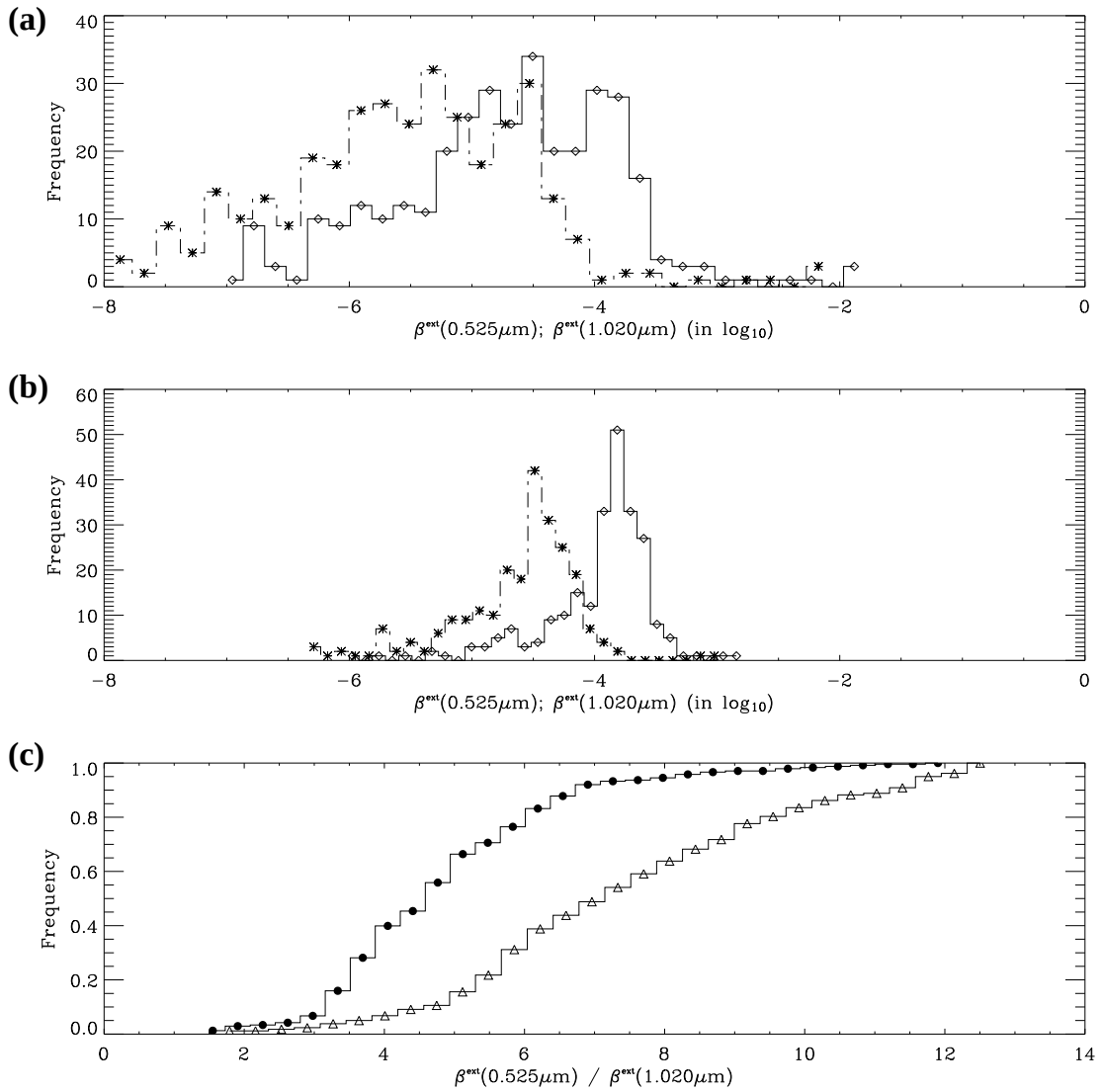

Figure 4.15: Histograms of (a) monomodal and (b) bimodal extinction at 0.525 μm (diamonds) and at 1.020 μm (asterisk). In (c) the extinction ratios at those two wavelengths are presented for monomodal (triangle) and bimodal (dots) aerosols, in the form of a normalised cumulative histogram.

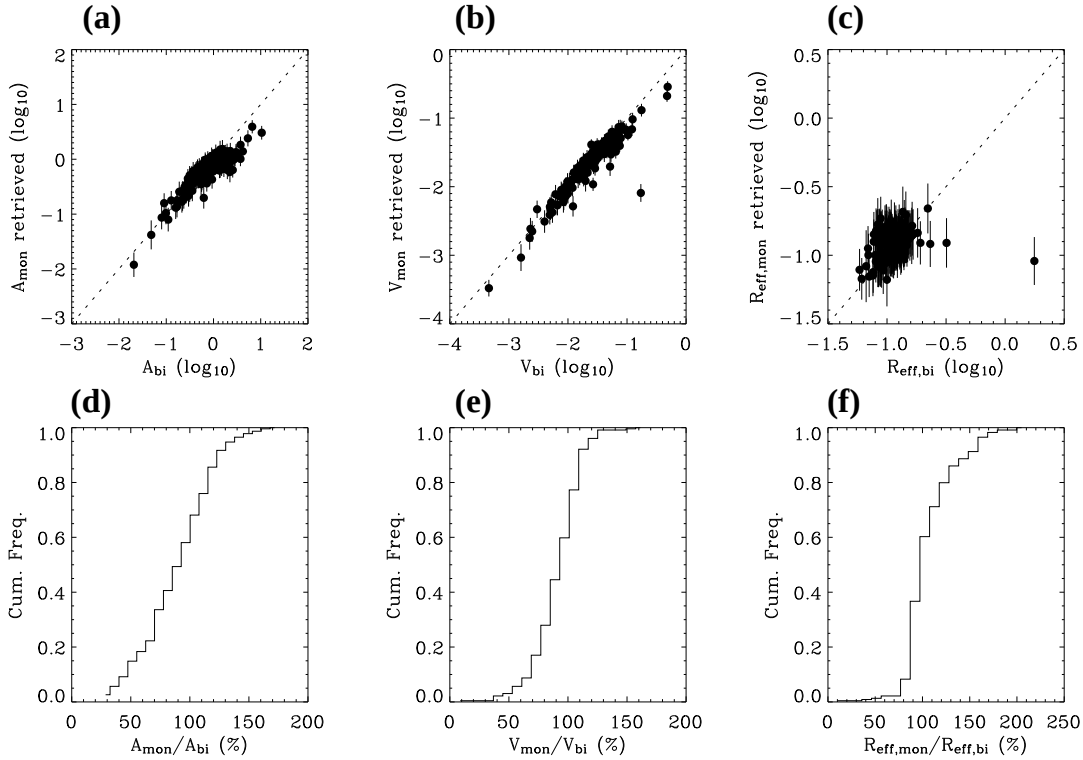

Figure 4.16: Maximum Noise Scenario: Monomodal parameters as retrieved from synthetic bimodal aerosol extinction is compared to the correct bimodal surface area density (a), volume densities (b), and effective radius (c). The frames (d-f) present the ratios of the retrieved and true quantities in the form of cumulative histograms. The correlation coefficients are 0.89 (A), 0.94 (V), and 0.26 (R_{eff}).

retrieved (one sigma) uncertainties confirms the generally good agreement between the data: it is 75% of all surface areas, 80% of all volume densities, and 86% of all resulting effective radii. From a statistical point of view, the retrieved results can be judged in very good agreement with the true solutions if 68% of all retrieved results agree with the true values within the error bars. In our data set, the observed agreement is higher implying that the retrieved uncertainties overestimate the true errors.

In the Minimum Noise Scenario (results not shown), the linear correlation coefficients between the retrieved monomodal and the correct bimodal values are 0.88 in A, 0.95 in V, and 0.45 in R_{eff}, all highly significant.

These results suggest that a distinct bimodal error cannot be detected. This agrees with investigations by Steele and Turco (1997) who found that it is possible for bimodal size distributions to account for extinctions generated from unimodal distributions and vice versa.

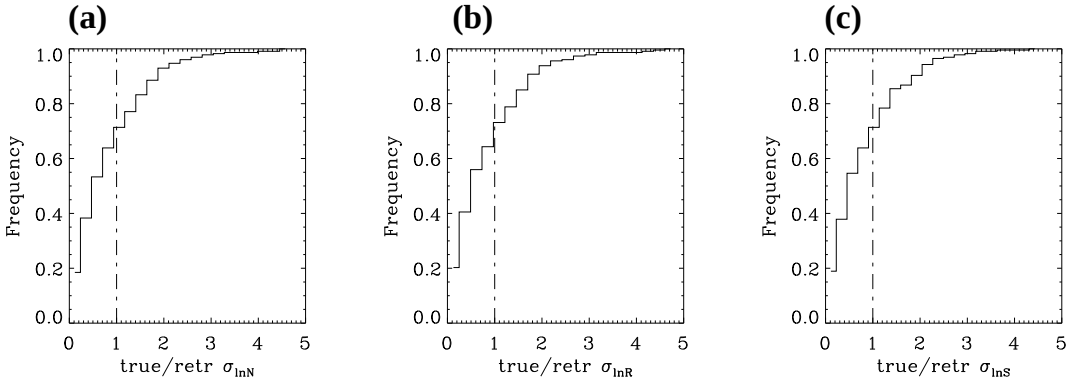

Figure 4.17: With minimum noise: Cumulative frequency distribution of the ratio of true and retrieved errors for (a) number density, (b) median particle radius, and (c) distribution width. The true error equals the retrieved error where $x = 1$, which is marked by the vertical broken line.

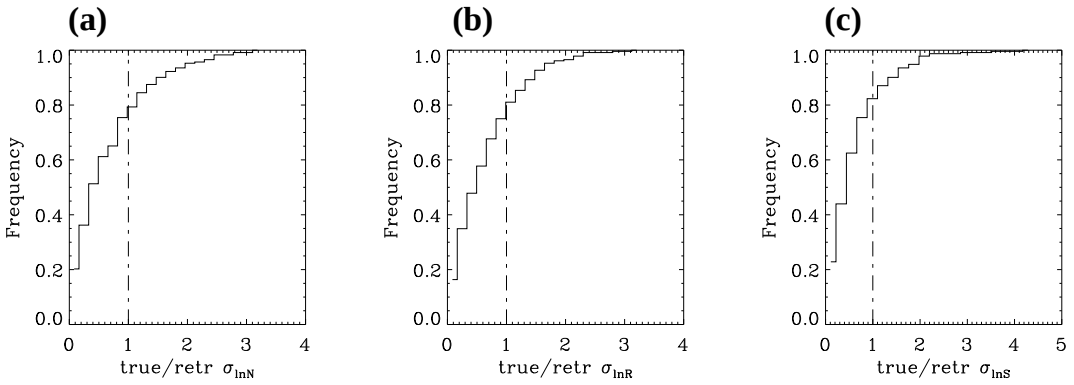

Figure 4.18: Same as Fig. 4.17, but for the Maximum Noise Scenario.

4.4.4 Retrieved Uncertainties versus True Errors

In order to validate the retrieved errors they can be compared to the true errors, which are the difference between the retrieved and the correct aerosol properties, $|\hat{x}_i - x_i|$. Since the retrieved error, $\sqrt{\hat{S}_{ii}}$, is the (one sigma) width of a Gaussian probability density function, we expect the true error to be smaller than the retrieved uncertainty in approximately 68% of all cases and larger in about 32% of all retrieval results. Figures 4.17 and 4.18 present the ratios of the true and retrieved errors in the form of cumulative histograms. A ratio < 1 (> 1) means that the retrieved uncertainty overestimates (underestimates) the true error. Table 4.7 summarizes which fractions (in %) of all 227 (minNS) and 232 (maxNS) results over- and under-estimate the true errors.

With very small experimental noise (Fig. 4.17, minNS), 63-70% of all retrieved uncer-

Over-/Underestimated (in %)	minNS	maxNS
$\ln N$	64/36	75/25
$\ln R$	64/36	75/25
$\ln S$	70/30	78/24
$\ln A$	63/37	82/18
$\ln V$	65/35	82/18
$\ln R_{\text{eff}}$	65/35	78/22

Table 4.7: Fractions (in %) of over- and underestimated retrieved errors.

tainties are observed to overestimate, and 30-37% to underestimate the true error. These numbers are very close to the expected 68 and 32% Gaussian fractions, which indicates that the retrieved uncertainties are a good representation of the true errors. With large experimental noise (Fig. 4.18, maxNS) the retrieval error overestimates the true error in 75-82% of all cases and under-estimates it in 18-25% of all cases. This indicates that large measurement uncertainties tend to overestimate the error. These results suggest that the Optimal Estimation algorithm produces uncertainty estimates which tend to be very good in the case of small extinction uncertainty (minNS), and conservative in the case of large measurement noise (maxNS). Where the retrieved uncertainty underestimates the true error, the two values generally do not differ by more than a factor 2.

4.4.5 Summary

The main results of this error analysis are as follows:

- The retrieval bias is always zero within rounding error, that is smaller than 10^{-6}, which means that the sensitivity of the retrieval method to any inverse model parameters (other than the *a priori* state) is zero.

- The retrieval solutions are partly determined by the information contained in the measurements and partly by the *a priori* information (constraint). The fractional contributions to the solution by the *a priori* constraint (as determined from the averaging kernels) is increasing with the measurement noise and was estimated to be an average 48% (minNS) or 76% (maxNS) in the retrieved number density, 23% (minNS, maxNS) in median particle radius, and 3% (minNS) or 70% (maxNS) in the retrieved distribution width.

- The retrieved uncertainties are mostly determined by the *a priori* uncertainty (smoothing error), whereas the contribution from extinction uncertainty (retrieval noise) tends to be small, that is between 5% (maxNS) and 30% (minNS). The influence of

the smoothing error component on $\hat{S}$ is observed to grow with increasing extinction uncertainty as the influence of the *a priori* constraint becomes more important.

- The numerical forward model error is estimated to be smaller than 1% in aerosol extinction. This is clearly smaller than the 10-60% measurement noise typically observed at 0.368 μm, but not automatically negligible compared to the 1-10% measurement noise typically observed at 1.020 μm.

- The forward model parameter error is estimated to be smaller than 3% in state space (aerosol properties). This is small compared to the retrieved uncertainties which are typically a factor of 20 larger.

- Under background conditions spectral extinction caused by bimodal aerosols cannot be distinguished from monomodal data because the orders of magnitude observed in the respective extinction values or in extinction ratios are basically identical.

- A comparison between the monomodal aerosol properties (as retrieved from bimodal extinction data) and the correct bimodal values suggest that errors in the retrieved integrated aerosol properties due to bimodal error tend to be negligible. This agrees with results by Steele and Turco (1997) who found that it is possible for bimodal size distributions to account for extinctions generated from monomodal distributions and vice versa.

- The retrieved errors were shown to be generally a good representation of the true errors, particularly with little noise (minNS). In the presence of large measurement noise (maxNS), the retrieved uncertainties tend to be conservative, that is slightly overestimating the true errors.

It is therefore concluded that under aerosol background conditions the Optimal Estimation algorithm achieves very good (minNS) or slightly conservative uncertainty estimates (maxNS), and that other additional error components (forward model errors, forward model parameter errors and bimodal errors) are considerably smaller.

4.5 Vertical Profiles

Thus far we have looked at the entire ensemble of retrieved results, regardless of measurement time or location. Presenting the results in the form of vertical profiles will highlight regions where the retrieved results tend to differ from the correct solutions. This information can give clues as to why some processes fail to converge toward the correct solution. For this analysis, the four profiles with the largest vertical coverage were extracted from the test data set. The associated measurement dates are May 1 and 28, June 27, and July 16 in 1999. Below, the retrieved results are presented together with the correct values, variable by variable. When interpreting the differences it is useful to remember that each profile was retrieved one height at a time starting at the lowermost height. Each (converged) retrieval solution then serves as the first guess state in the following retrieval process. The uppermost retrieval solution serves as first guess value in retrieving from the lowermost measurement of the following profile. The relative difference between any two profiles at one particular height i is given by

$$\Delta x_i = \frac{(x_{\text{true,i}} - \hat{x}_i)}{x_{\text{true,i}}} * 100, \tag{4.14}$$

where $\hat{x}$ is the retrieved state value and Δx is given in % of the correct value. The differences are calculated at each height and do not take into account the retrieved uncertainties. From all differences within a profile, the profile mean difference is computed as the arithmetic mean of the absolute values of all N relative differences within a profile

$$\bar{\Delta x} = \frac{1}{N} \sum_{i=1}^{N} |\Delta x_i|. \tag{4.15}$$

Comparisons are shown for one noise scenario only (maxNS), as the results in the other noise scenario (minNS) are qualitatively similar in the following aspects. Good/poor agreement between the retrieved and the correct results is generally observed at the same locations. At those altitudes where the largest differences occur, the absolute differences tend to be smaller in the minNS than in the maxNS, but they may still be larger than the retrieved uncertainties (which are generally smaller in the minNS than in the maxNS). The relative differences between any two profiles are observed to be generally unbiased, regardless of the experimental noise level.

Figures 4.19.a/c/e/g present the profiles of number density. The true profiles show that there is a general decrease of number density with height. Between 20 and 35 km the gradient is generally small. The retrieved N is close to the first guess value (solution of the level below) and also to the *a priori* value (which is marked by the vertical dashed line), and the retrieved number densities tend to agree with the correct values within

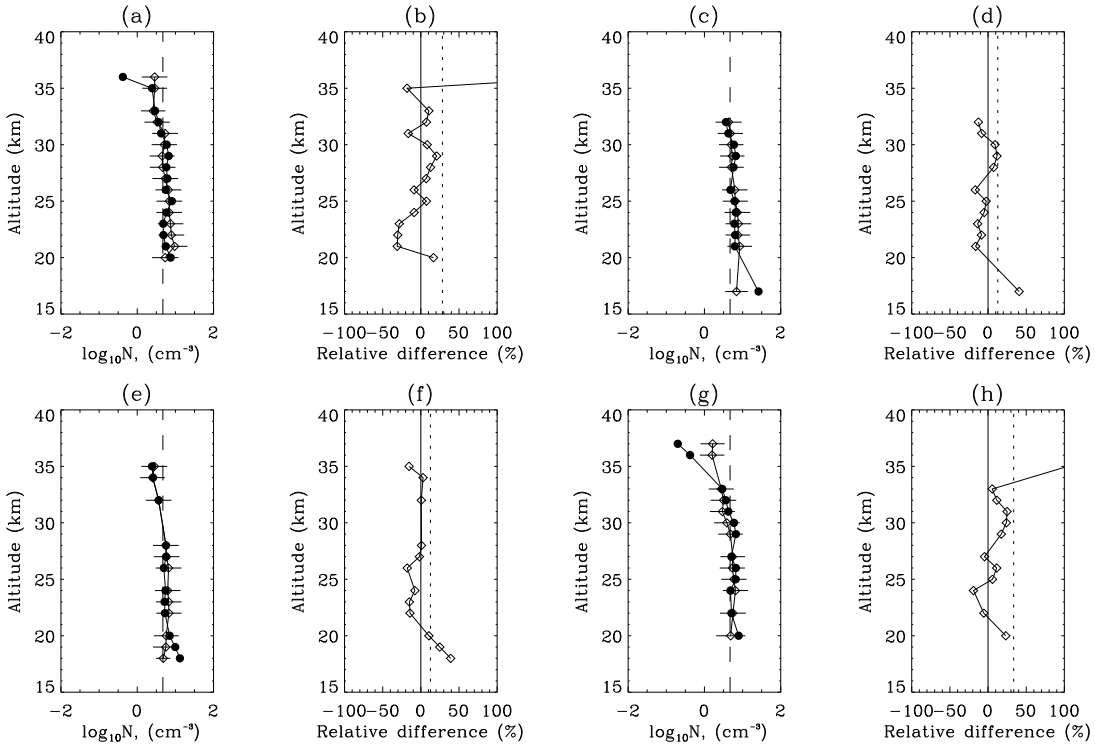

Figure 4.19: Maximum Noise Scenario: Profiles of retrieved (diamonds) and true (dots) total number density N in cm^{-3} (a, c, e, g) and the associated differences (in %) to the true solution (b, d, f, h). The vertical lines indicate the *a priori* mean state (dashed), the zero difference (solid), and the mean profile difference (dotted).

the uncertainties. Above 35 km and below 20 km, the vertical gradients as measured *in situ* are stronger. Differences in N of more than an order of magnitude occur between adjacent altitude levels and the distance to the *a priori* mean increases. In those regions, the retrieved profiles tend to differ from the correct values. The observation that in these locations the retrieved N are generally closer to the *a priori* value than the correct number densities is most likely a consequence of the small information content (with respect to number density) and the associated stronger influence of the *a priori* constraint. The relative differences (Fig. 4.19.b/d/f/h) are generally unbiased, and the average deviations from the correct profiles are observed to range between 12% and 34%.

Figure 4.20 presents the profiles of median particle radius. Generally, R is well retrieved and agrees with the correct solutions within the error bars. The two profiles differ only in two locations (Fig. 4.20.a at 36 km, and Fig. 4.20.c at 16 km). From 35 to 26 km in Figure 4.20.a the retrieval is likely to have converged to a local minimum rather than following the stronger gradient toward larger particles sizes as shown by the *in situ* profile.

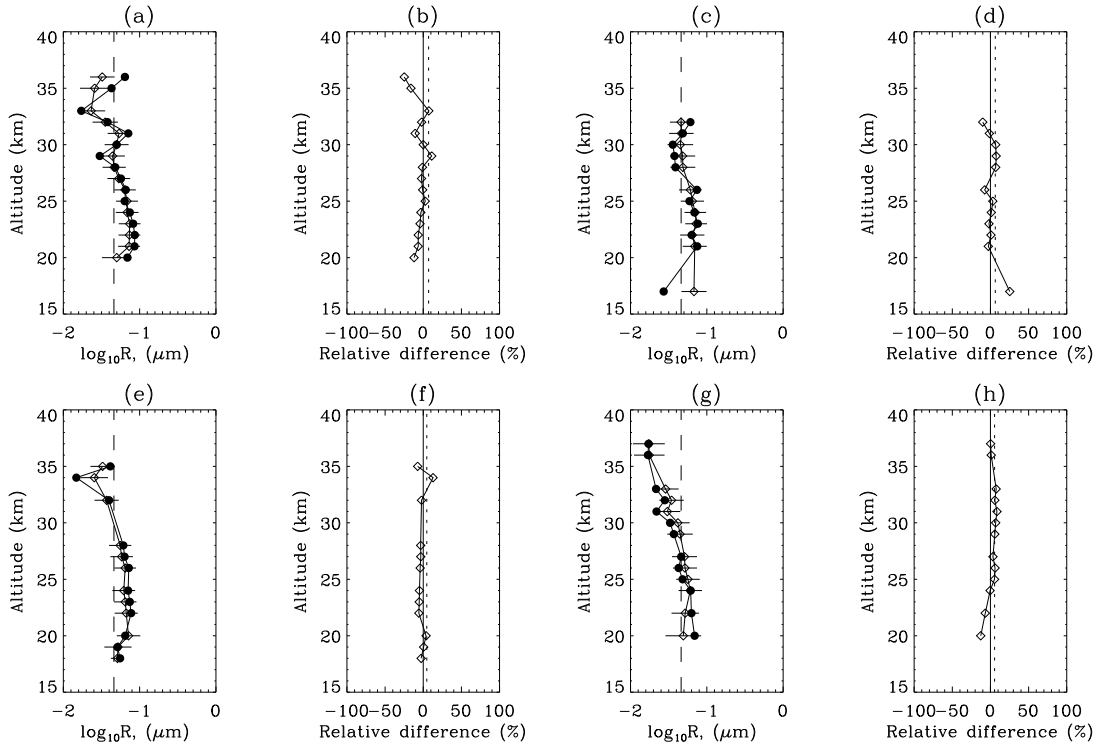

Figure 4.20: Maximum Noise Scenario: As Fig. 4.19 but for median radius R.

Interestingly, at 17 km in Fig. 4.20.c the retrieval converged to a solution which is farther away from the first guess than the correct solution, which is very close to the first guess. This shows that a good first guess in one state variable is no guarantee for convergence to the correct solution, possibly because other variables are being optimized at the same time. In Figure 4.20.g at the upper end of the profile the retrieved solutions coincide with the correct solutions, although the particle size decreases and the difference to the *a priori* mean increases. This indicates that the *a priori* constraint is 'loose' enough not to force all solutions close to the *a priori* mean. The relative differences between the retrieved and the correct profile in median radius are generally observed to be unbiased. The profile mean differences tend to be smaller than those in number density; they are observed to range between 4% and 7%. Comparison of the particle median size with number density (Fig. 4.19) shows that an overestimate in N is associated with an underestimate in R (Fig. 4.20.a at 36 km) and vice versa (Fig. 4.20.c at 17 km). This makes sense as a few large particles can be expected to produce the same extinction effect as many small particles.

Figure 4.21 presents the profiles of lognormal distribution width. The retrieved S tends to agree with the correct solutions within the error bars except near the upper ends of

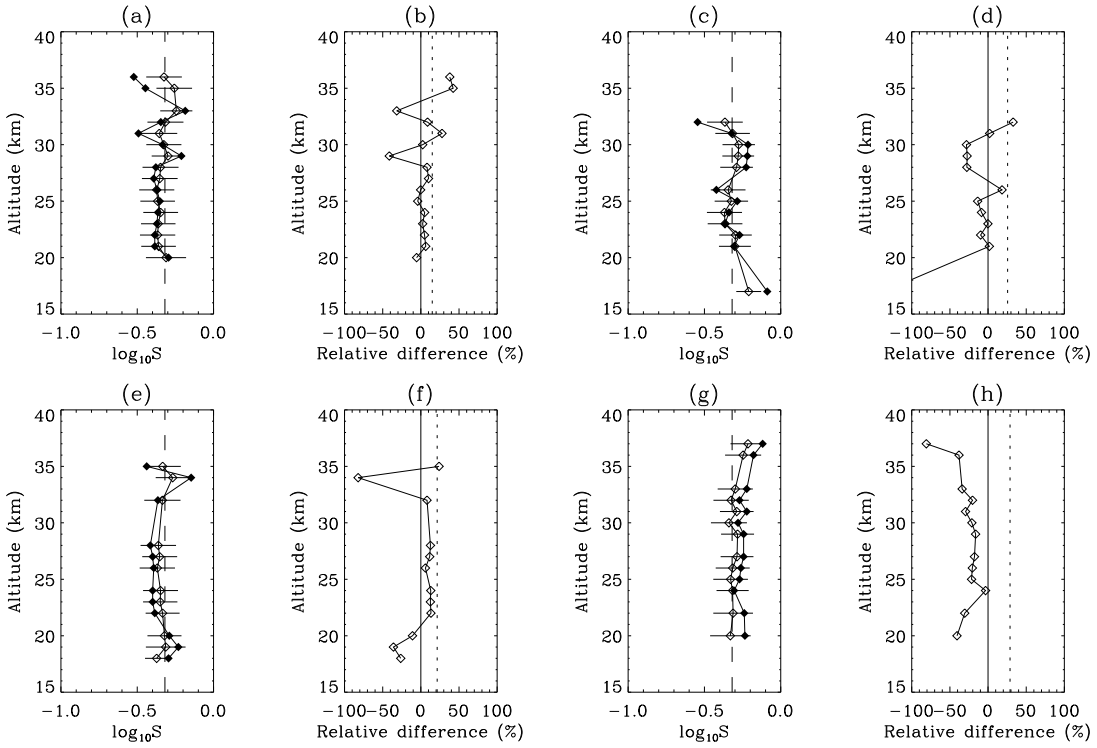

Figure 4.21: Maximum Noise Scenario: As Fig. 4.19 but for distribution width S.

Figures 4.21.a/c and near the lower end of Fig. 4.21.c. In all three cases the retrieved solutions do not follow the strong gradient of the correct profiles, but converge to local minima that are closer to both the first guess and the *a priori* mean values than the correct solution. The average profile differences are observed to range between 15 and 29%.

Figure 4.22 presents the profiles of surface area density as derived from the size distribution parameters described above. The retrieved profiles are observed to match the correct ones within the error bars except at and below 20 km (Fig. 4.22.a/e/g), where the retrieval converged to local minima that are close to the respective first guess values and to the *a priori* mean, therefore underestimating the correct solutions. It is interesting to note that a good match between the retrieved and the correct A can be observed at 17 km in Figure 4.22.c, although all three state variables (N, R, and S) differed from the correct solutions. The relative differences (Fig. 4.22.b/d/f/h) are observed to be generally unbiased, and the profile mean differences range between 11% and 71%.

Figure 4.23 shows the profiles of volume density, which in their vertical structure are observed to be very similar to surface area density. In general, there is a good match between the retrieved and the true solutions except at low altitudes, where the retrieved volumes tend to be too small. The profiles of relative differences in V are almost identical

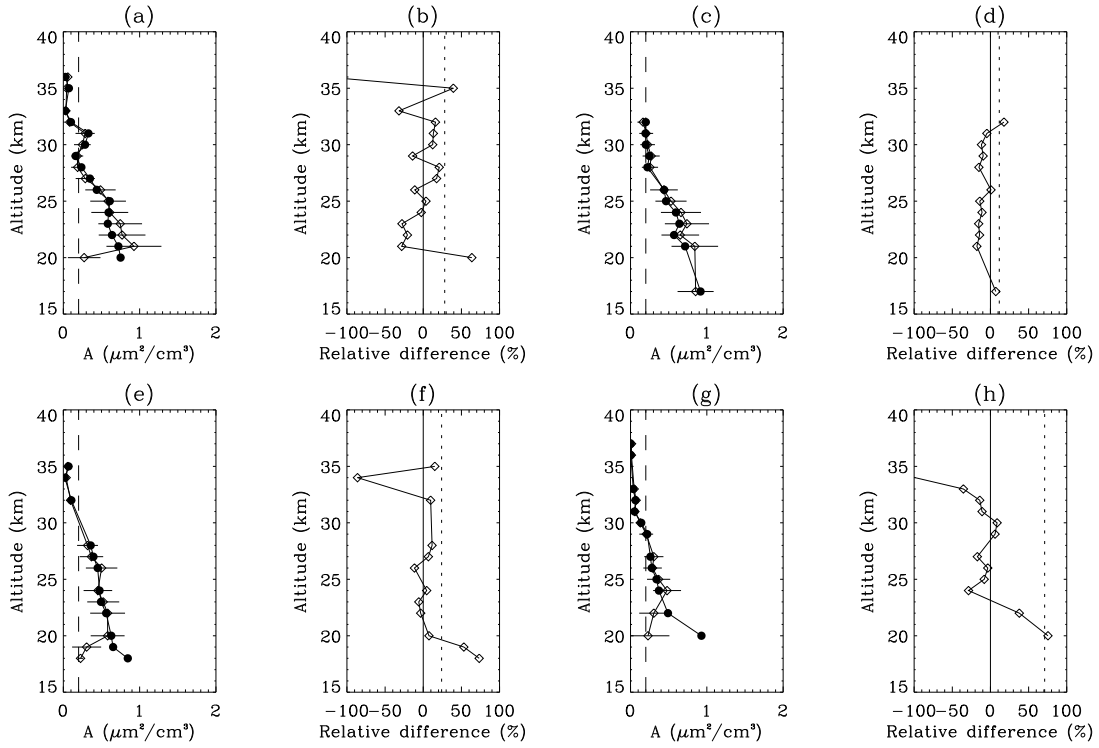

Figure 4.22: Maximum Noise Scenario: As Fig. 4.19 but for surface area density A.

with those in A, and the observed average differences per profile range between 10% and 51%.

Figure 4.24 presents profiles of effective particle radius, which is basically the area weighted volume and derived from the above surface areas and volume densities. As expected (given the good results in A and V), the retrieved and true profiles of R_{eff}mostly agree. Significant differences are observed at the uppermost (Fig. 4.24.a/g) and lowermost heights (Fig. 4.24.e/g), where the retrieval converged to a local minimum near the first guess. The observed profile mean differences range between 4% and 15% and hence are smaller than those in A and V, indicating a generally better fit.

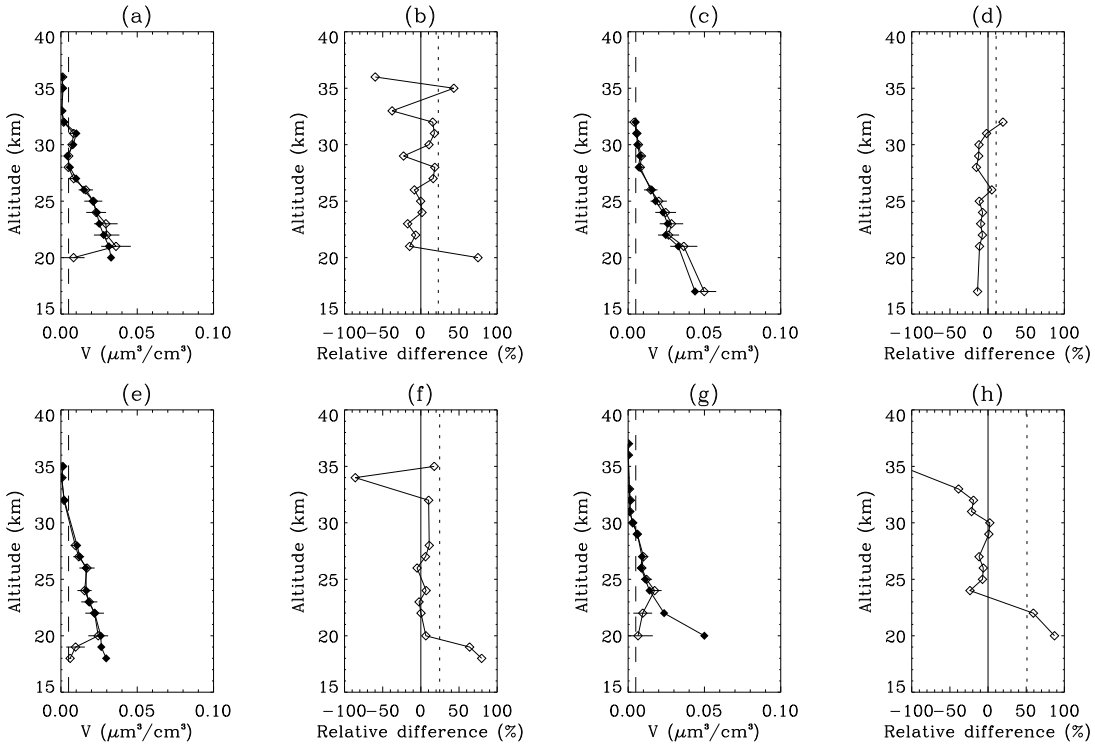

Figure 4.23: Maximum Noise Scenario: As Fig. 4.19 but for volume density V.

Summary

The comparison between the true and the retrieved aerosol properties in the form of vertical profiles has shown:

- Discrepancies are mainly observed at the lower and upper ends of the profiles, that is at or below 20 km and at 36 km and above, where the differences between the correct solutions and the *a priori* values tend to be large. These discrepancies are likely to be caused by local cost function minima near the first guess values, which are closer to the *a priori* mean than the correct solution.

- In general a good first guess is helpful, however it does not guarantee an accurate retrieval solution. It is possible for a retrieval to converge to a solution which is further away from the first guess than the correct solution, although the latter happened to be situated very close to the first guess value (Fig. 4.20.a/c).

- An observed underestimation in the number density tends to be correlated with an overestimated particle median radius. This can be explained by a few large particles being able to produce the same extinction as many small particles.

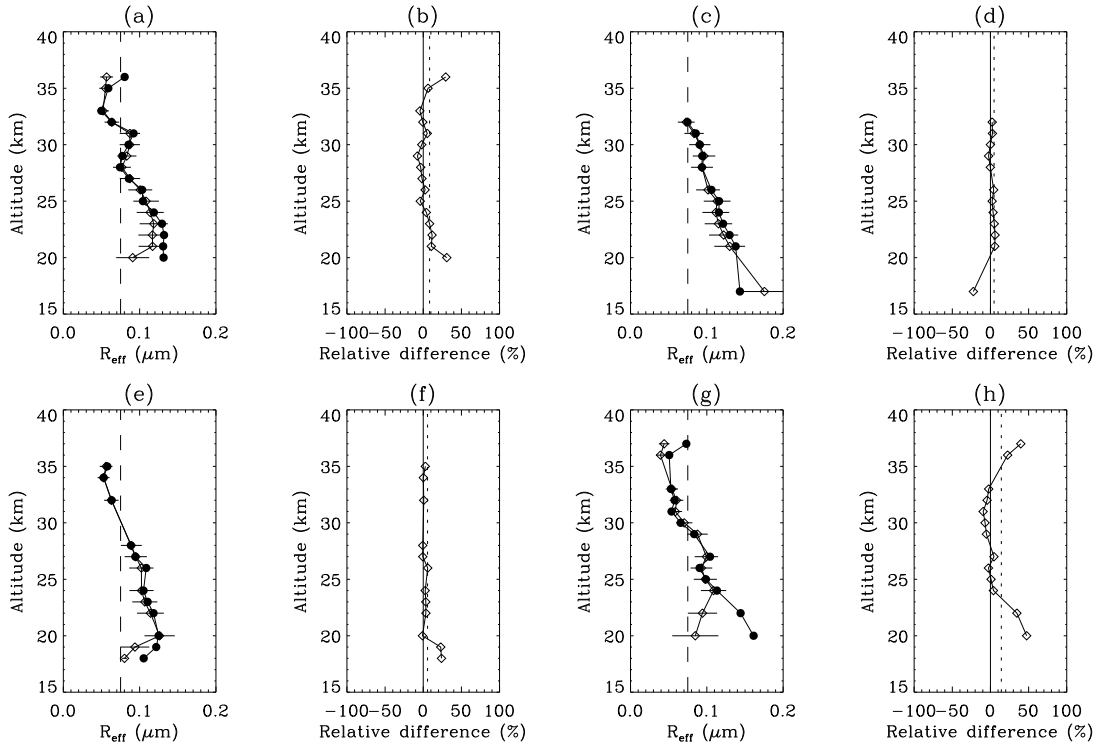

Figure 4.24: Maximum Noise Scenario: As Fig. 4.19 but for effective particle radius R_{eff}.

- It is possible for the derived integrated properties to be close to the correct solution, although N, R, and S differ from the correct solutions.
- The profile mean differences range between 12-34% in number density, 4-7% in median radius, 15-29% in lognormal distribution width, 11-71% in surface area density, 10-51% in volume density, and 4-15% in effective radius (in the realistic noise scenario).

All in all, the new algorithm was shown to be capable of retrieving aerosol microphysical properties well, even under relatively strong noise conditions.

Chapter 5

Retrieval from SAGE II Data

In this chapter, the new Optimal Estimation retrieval algorithm is applied to SAGE II satellite measurements of aerosol extinction made in 1999. This period was selected because these measurements are part of the longest volcanically quiescent period in the record which began around 1997. The aerosols can therefore be assumed to consist of sulphuric acid droplets with a monomodal size distribution. The objective of this chapter is to assess the performance of the new Optimal Estimation retrieval algorithm in a real-life situation.

In Section 5.1, the retrieved aerosol properties and associated uncertainties are presented and compared to the *a priori* distributions and to the uncertainties estimated by other researchers. In Section 5.2, a small number of vertical profiles (extracted from the above data set) are compared to correlative profiles measured *in situ* in order to estimate the plausibility of the retrieved aerosol properties. In Section 5.3, the Optimal Estimation retrieval results are compared with aerosol properties as retrieved by NASA from the same SAGE II extinction data but using the Principal Component Analysis approach. In Section 5.4 the Optimal Estimation results are used to generate monthly latitude-altitude cross sections of zonally averaged aerosol properties. The quality of these monthly means is assessed with respect to their suitability for examining seasonal change.

Month	Mar	Apr	Jun	Jul	Sept	Dec
Analysed	18,717	17,601	19,722	11,680	18,238	19,778
Converged (%)	99.9	99.9	99.9	99.9	99.9	99.9
Accepted (%)	90.1	91.4	91.7	89.5	90.1	90.1

Table 5.1: Total number of retrievals, fraction of all converged retrievals (in %) and accepted solutions (in %). The criteria are described in Section 4.3.

The retrieval algorithm was applied to aerosol extinction measured during six months with approximately 11,600-20,000 monthly measurements. These measurements are independent of the *a priori* data used in the retrieval algorithm. Aerosol composition (concentration) and refractive indices are calculated as described in Section 2.2, based on the prevailing ambient temperature and humidity and using laboratory data by Steele and Hamill (1981) and Russell and Hamill (1984) (to determine the aerosol acidity) and a model by Semmler *et al.* (2003) (to determine the associated refractive index). Table 5.1 lists the number of monthly measurements analysed, the percentage of those that converged and those that pass the quality filter tests identified in Section 4.2. It can be observed that the retrieval process converged nearly every time (99.9%), and that more than 90% of the achieved results passed the quality screening.

5.1 Retrieved Aerosol Properties

The month with the largest number of measurements, that is December 1999, will now be used to look at the results in more detail. Measurements with uncertainties larger than 99% are discarded *a priori* and not used in the following data analysis. Figure 5.1 presents the aerosol number densities, median radii and distribution widths as retrieved from SAGE II measurements in December 1999. It can be observed that the retrieved aerosol properties are of the same order of magnitude as the *a priori* distributions. Comparison with the *a priori* mean and standard deviation (as marked by the full and dash-dot vertical lines respectively), however, shows that the retrieved values are shifted toward higher values, but with the majority of all values still lying between the *a priori* mean state and the *a priori* state plus or minus one standard deviation.

Figure 5.2 presents the associated uncertainties. These are typically 60-75% in number density, 30-40% in median radius, and 19-20% in distribution width. In comparison, the *a priori* uncertainties on N, R, and S were respectively 93, 61, and 31% (Sec. 3.2), which means that analyzing the measurements has improved our *a priori* knowledge about the size distribution parameters (uncertainties) by an average 26% (N), 46% (R), and 55% (R_{eff}). The *a priori* and retrieved ensemble mean size distribution parameters and associated uncertainties are listed in Table 5.2.

Figure 5.3 presents the uncertainties in the derived integrated properties. These are typically of the order of 20-30% in surface area density, 5-20% in volume density, and 10-15% in effective radius. In comparison, the *a priori* uncertainties on A, V, and R_{eff} were respectively 146, 12, and 13%, which means that analyzing the measurements has improved our *a priori* knowledge about the integrated aerosol properties (or the uncertainties) by approximately 84% (A), 93% (V), and 68% (R_{eff}). The retrieved ensemble mean integrated aerosol properties and associated uncertainties are listed in Table 5.3.

Table 5.4 lists the OE uncertainties together with results reported by other researchers

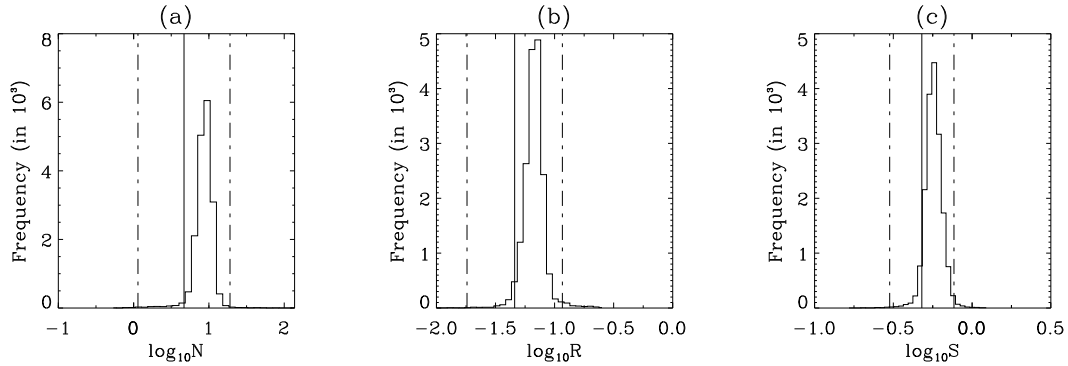

Figure 5.1: Histograms of number density (a), median radius (b), distribution width (c) as retrieved from SAGE II measurements of aerosol extinction in December 1999. The vertical lines indicate the *a priori* state (solid), and the *a priori* state plus or minus one standard deviation (dash-dot).

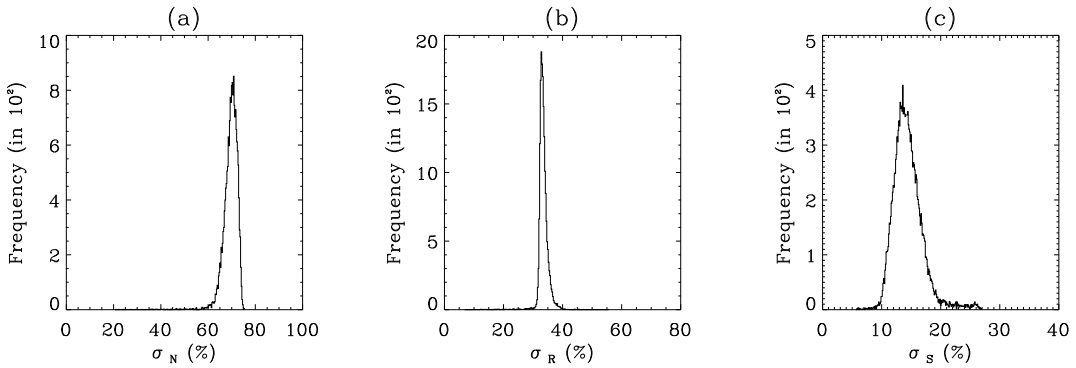

Figure 5.2: SAGE II, 12/1999: Histograms of the retrieved uncertainties (in %) in number density N, median particle radius R, and distribution width S.

(Sec. 2.4). When comparing the different uncertainties it has to be kept in mind that these values are all based on different data sets, were retrieved using different methods, and may be estimates of partial errors only (e.g. Tab. 5.4, source (2) and (4)) rather than a realistic estimate of the total difference between the retrieved and the true solutions. Nevertheless, the table gives an impression of what has been achieved and reported in the literature in this area of research. Generally, it can be observed that the orders of magnitude of the OE errors in surface area, volume density and effective radius agree with what has been reported by other researchers. Considering that the uncertainties in A as reported by Steele *et al.* (1999) and Steele and Turco (1997) are expected to be higher in reality, due to disregarded systematic (method bias) error and contribution from particles smaller than 0.1 μm, respectively, the OE uncertainties in A, V, and R_{eff} are among

	SAGE II, Dec 1999
	N, R, S
Ensemble Mean	9.0, 0.069, 0.57
A priori	4.7, 0.046, 0.48
	σ_N, σ_R, σ_S (%)
Ensemble Mean	69, 33, 14
A priori	93, 61, 31

Table 5.2: Ensemble mean retrieved size distribution parameters with associated uncertainties (in %). Number density is given in cm^{-3}, median radius in μm, and lognormal distribution (half) width in log of μm.

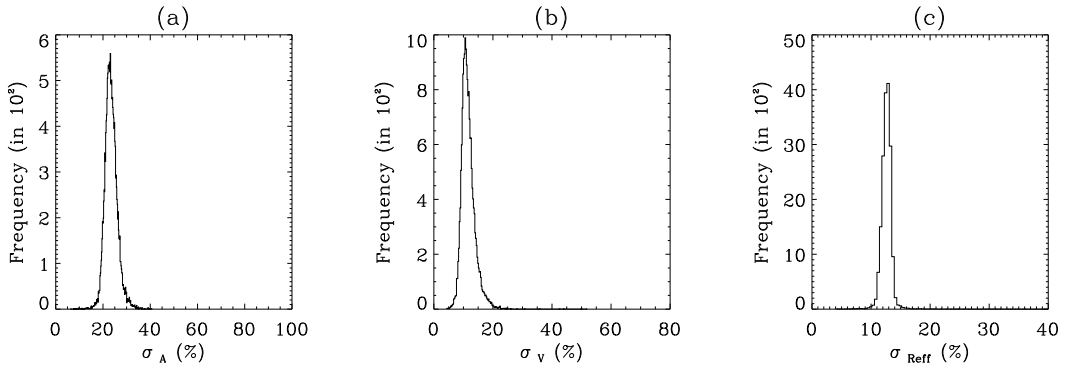

Figure 5.3: SAGE II, 12/1999: Uncertainties in surface area, volume and effective radius as derived from the retrieved size distribution parameters.

the smallest. In number density, the uncertainties reported by Bingen *et al.* (2004b) are on average larger than the OE error estimates, whereas those estimated by Wang *et al.* (1989) appear considerably smaller. However, the reported 11% only consider particles greater than 0.15 μm, whereas the great majority of the retrieved sizes is smaller (Fig. 5.1). Similarly, uncertainties in median particle radius estimated by Wang *et al.* (1989) are also smaller than those reported by Bingen *et al.* (2004b) or those achieved using OE (this work), but only particles larger than 0.1 μm are considered in their calculations. Although particles smaller than 0.1 μm contribute little to the total aerosol extinction (at visible wavelengths), their contribution is important to get accurate estimates of the retrieved aerosol properties (Sec. 1.2). In distribution width, the OE uncertainties are an order of magnitude smaller than those estimated by Bingen *et al.* (2004).

This comparison suggests that the OE retrieval results tend to be at least as good as those results achieved by other researchers using different methods. In Section 5.3 a

	SAGE II, Dec 1999
	A, V, R_{eff}
Ensemble Mean	1.00, 0.05, 0.16
A priori	0.20, 0.005, 0.075
	σ_A, σ_V, σ_{Reff} (%)
Ensemble Mean	23, 12, 13
A priori	146, 179, 40

Table 5.3: Ensemble mean retrieved surface area density, volume density, and effective radius with associated uncertainties (in %). Surface area density is given in $\mu m^2 cm^{-3}$, volume density in $\mu m^3 cm^{-3}$, and effective radius in μm.

Source / Method	σ_N (in %)	σ_R (in %)	σ_S (in %)	σ_A (in %)	σ_V (in %)	σ_{Reff} (in %)
(1) / OE	60-75	30-40	10-20	20-30	5-20	10-15
(2) / PCA				(15-20)+50		
(3) / PCA				30	12-25	
(4) / CLI				25+	15+	15+
(5) / RMST				8-50	5-25	6-36
(6) / RIM	50-200	35-50	100-250			
(7) / NIM	<11	5-28				
(8) / EM				<30	<15	<15

Table 5.4: Overview of retrieved errors (in %) that were achieved by different retrieval methods. The '+' indicates that the true error is higher due to other disregarded uncertainty components. The methods and the conditions under which these uncertainties were achieved were described in Section 2.4. The data sources are: (1) SAGE II, Dec. 1999 (this thesis), (2) Steele *et al.* (1999), (3) Thomas and Poole (1993), (4) Steele and Turco (1997), (5) Anderson *et al.* (2000), (6) Bingen *et al.* (2004b), (7) Wang *et al.* (1989), and (8) Hervig *et al.* (1998). The acronyms stand for Principal Component Analysis (PCA), Constrained Linear Inversion (CLI), Randomized Minimization Search Technique (RMST), Regularized Inversion Method (RIM), Nonlinear Iterative Method (NIM), and Empirical Method (EM).

further comparison will be conducted between the new OE retrieval results and aerosol integrated properties as retrieved by NASA LaRC applying the Principal Component Analysis approach to the same data set (SAGE II data, in December 1999).

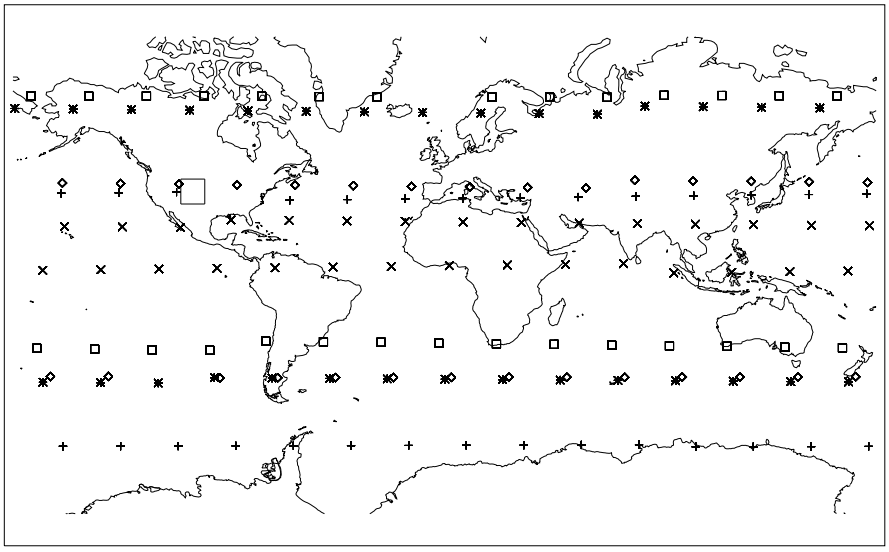

Figure 5.4: Locations of SAGE II measurements taken on the following days in 1999: March 18 (+), April 20 (∗), June 23 (◇), September 16 (□), December 10 (X). On those days, *in situ* measurements exist that were measured by the University of Wyoming over Laramie (41°N, 105°W, indicated by the large square).

5.2 Comparing OE Profiles with Correlative in situ Measurements

A widely used way of validating remotely sensed information is to compare it to correlative measurements. Such comparisons cannot be conclusive but give indications whether or not the retrieved results are plausible. Data sets obtained from different instruments can always be expected to differ in their details because, naturally, the measurement volumes are not identical. Here we will use correlative *in situ* measurements to validate the retrieved OE results. The two data sets are expected to differ mainly for the following reasons:

1. The measurement locations/volumes of remote sensing data and *in situ* measurement techniques are different: The *in situ* profiles reflect the state of the atmosphere in an approximately vertical column of the atmosphere. In time the measurements at different altitude levels are only minutes apart. In contrast, signals measured remotely via satellite are integrated data which means that all contributions along the line-of-sight of the instrument are added and allocated to a particular tangent height. In addition, the measurement geometry of the solar occultation technique (Fig. 2.5) also causes all tangent points within a profile to be horizontally shifted against each other. This means that the SAGE II data give averaged information over

a relatively large region while *in situ* instruments yield more localized information. Consequently, vertical profiles of aerosol properties as measured by the SAGE II instrument are likely to be smoother than the *in situ* profiles.

2. It is usually not possible to find correlative measurements that match in measurement time as well as location. At best, the remotely sensed and the *in situ* data are close in location and were measured not too long apart. Here, the measurement locations of the profiles to be compared differ by up to 5° in longitudinal and 2° in latitudinal direction. The measurement dates differ by up to 6 days because of a lack of more similar data sets (with respect to measurement time and location).

3. The atmosphere is constantly changing, which can increase the discrepancies between different data sets that were not measured at exactly the same time and location. However, the importance of this effect depends on the location (and the measurement date). This is because in the longitudinal direction and away from any volcanic influence the atmosphere tends to be well mixed. Small differences in time and longitude are therefore not expected to affect the comparability of the profiles. Latitudinal differences in the measurements are more likely to cause discrepancies in the profiles to be compared because the global circulation patterns create stronger gradients in the north-south direction.

In 1999, *in situ* measurements of aerosol particle size distributions (as measured by the University of Wyoming Optical Particle Counters near Laramie/Wyoming) are available on six dates: March 22, April 20, June 23, July 21, September 16, and December 10. (These are independent of the data going into the *a priori* constraint (Sec. 3.2)). Figure 5.4 shows the locations of SAGE II measurements that were collected on those same days. It can be observed that only two SAGE II measurements are located near Laramie. In order to have a few more SAGE II measurements to compare, the time constraint was slightly relaxed to allow for data that were measured within a few days of the Laramie data. Table 5.5 lists the measurement times and locations of two *in situ* profiles, four profiles measured by SAGE II. The approximate temporal and spatial differences (with respect to the *in situ* measurements) between any two correlative profiles are up to approximately 6 days time wise, 2 degrees in latitude, and 5 degrees in longitude. Vertical profiles of retrieved aerosol properties are compared variable by variable. The relative difference between any two measurements at a particular altitude i is given by

$$\Delta x_i = \frac{(x_{i,\text{insitu}} - \hat{x}_i)}{x_{i,\text{insitu}}} * 100, \tag{5.1}$$

where $\hat{x}_i$ is the aerosol property as retrieved from SAGE II measurements, and Δx_i is given in % (of the *in situ* value). The profiles cover an altitude range between approximately 15 and 24 km. As the retrieved data are available on a 500 m vertical resolution grid and

Data	date (Δ in days)	lat (Δ in °)	lon (Δ in °)
in situ	23 Jun	41°N	105°W
	10 Dec	41°N	105°W
SAGE II	22 Jun (-1)	39°N (-2)	106°W (-1)
	23 Jun (0)	43°N (+2)	110°W (-4)
	14 Dec (+4)	39°N (-2)	100°W (+5)
	16 Dec (+6)	42°N (+1)	105°W (0)

Table 5.5: Dates and approximate locations of balloon borne *in situ* measurements (University of Wyoming) and correlative SAGE II data. The differences (Δ) between any two profiles that are being compared are given in brackets.

the *in situ* data on a 1 km resolution grid, differences are calculated on the coarser 1 km grid and at the same nominal heights. These differences can then be compared to both the retrieved and the *in situ* uncertainties reported by the University of Wyoming science team. Profile mean differences are calculated as the arithmetic mean of all (absolute) differences in the profile

$$\overline{\Delta x} = \frac{1}{N} \sum_{i=1}^{N} |\Delta x_i|. \tag{5.2}$$

Comparisons between vertical profiles of the retrieved and the *in situ* values are presented for those variables which are of direct interest with respect to the physical and chemical processes. These are the number density, median radius, surface area density, volume density, and effective radius. Vertical profiles of distribution width are only indirectly presented as they were used to calculate (the vertical profiles of) the integrated aerosol properties A, V, and R_{eff}.

Figure 5.5 presents the vertical distributions of number density. Generally, it can be observed that the *in situ* profiles cover a wider range of number densities than the retrieved profiles. Agreement within the OE error bars can be observed between 18 and 24 km in June (Figure 5.5.a/c) and between 16 and 27 km in December (Figure 5.5.e/g). Below 16 km the *in situ* number densities are generally larger than the retrieved solutions by up to an order of magnitude. The observed profile mean differences range between 20 and 36%.

Figure 5.6 presents the profiles of median particle radius. Generally, the retrieved and the monomodal *in situ* profiles are observed to agree within the errors (*in situ* and/or retrieved uncertainty estimates), except at the lowermost heights at and below approximately 15 km, where the *in situ* particle sizes decrease more rapidly (with decreasing altitude) than the

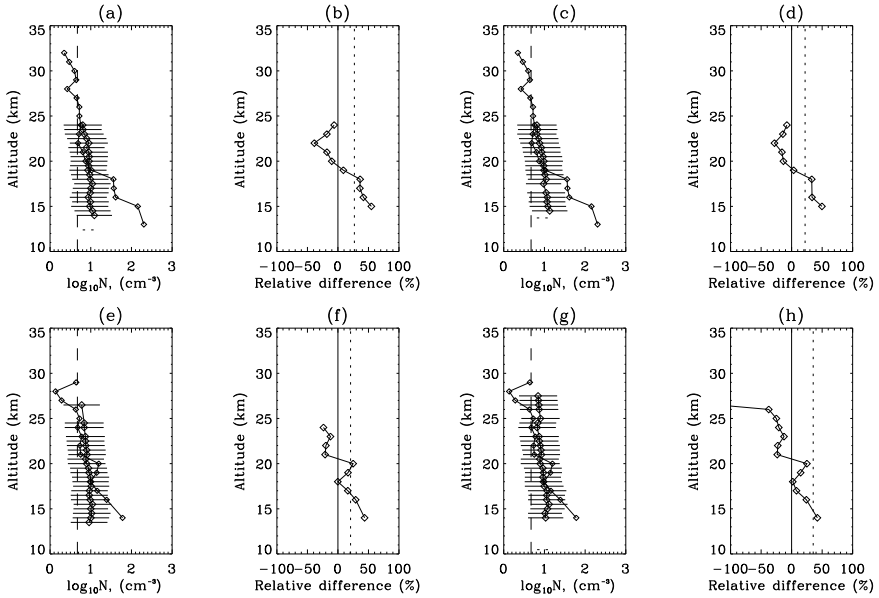

Figure 5.5: Vertical profiles of number density. Frames a/c/e/g: Retrieved (diamonds with error bars, measured on June 22 (a) and June 23 (c), 1999) and correlative *in situ* measurements (full dots, measured on June 23, 1999) of N; the vertical dashed line marks the *a priori* mean N, and the short horizontal dotted line marks the tropopause level (NMC data) at the time of the SAGE II measurements. Frames b/d/f/h: Associated relative differences. The long vertical lines mark the zero (solid) and the profile mean difference (dotted). The *in situ* uncertainties are approximately 10% [Deshler *et al., 2003*].

retrieved solutions. The average profile differences are observed to range between 9 and 13%.

When comparing the particle radii (Fig. 5.6) with number density (Fig. 5.5), it can be observed that where the retrieved number densities are smaller than the *in situ* values the associated particle median radii are larger. A similar observation was made in the validation chapter (Sec. 4.5) and can be explained by the connection between particle number, size and aerosol extinction: a few large particles can produce the same extinction as many small particles. It should also be noted that the strongest deviations in both N and R occur at altitudes where the respective *in situ* values are larger (N) or smaller (R) than the *a priori* mean plus or minus one *a priori* standard deviation (Fig. 5.1). Due to the *a priori* constraint it is rather unlikely to retrieve solutions as high (in N) or as low (in R).

Figure 5.7 shows the profiles of surface area density as derived from the retrieved size distribution parameters. Generally, it can be observed that the retrieved values follow closely the vertical structure of the *in situ* surface areas. In December (Fig. 5.7.e-h),

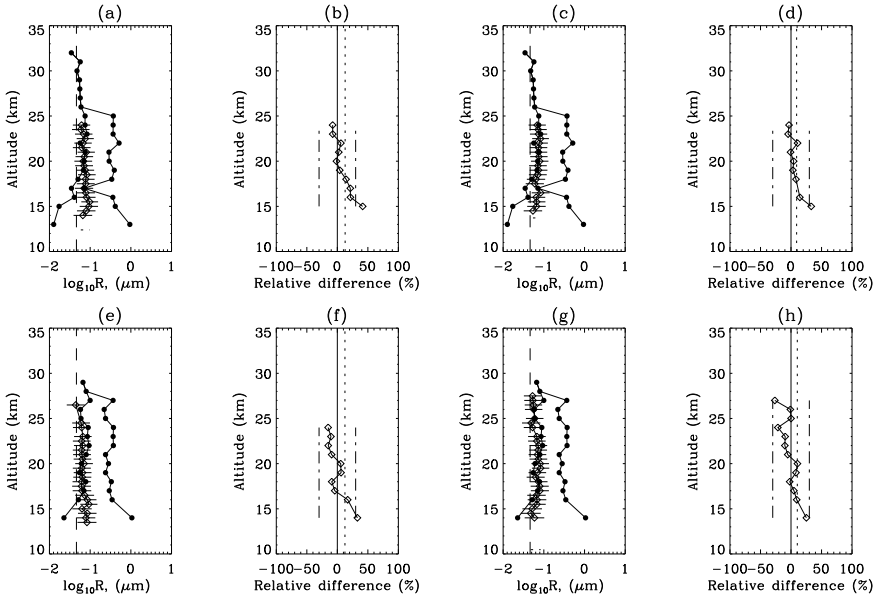

Figure 5.6: As Figure 5.5, but for median particle radius. The *in situ* data show two profiles, the smaller values being monomodal data, the larger values being the smaller mode of bimodal aerosols. Relative differences (frames b/d/f/h) are calculated with respect to the monomodal data, and the *in situ* uncertainties (as marked by the dash-dotted vertical lines) are 30% [Deshler *et al., 2003*].

the retrieved and the *in situ* profiles agree within the uncertainties except above 25 km (Fig. 5.7.g/h). In June (Fig. 5.7.a-d), agreement between the profiles (with respect to the *in situ* errors) is observed except between 20 and 23 km (Fig. 5.7.b/d), where the retrieved surface areas are larger and also further away from the *a priori* mean state than the *in situ* values. The observed profile mean differences range between 10% and 46%.

Profiles of aerosol volume density are presented in Figure 5.8. The vertical structure of these profiles is observed to be similar to the structure of surface area density, but while the surface area differences are unbiased, the relative differences in volume density (Fig. 5.8.b/d/f/h) are mostly negative, because the retrieved values tend to be larger than the *in situ* profiles. In June (Fig. 5.8.a-d), only little agreement (within the uncertainties) is observed. On December 14 (Fig. 5.8.e-f) the two profiles agree well above 15 km. On December 16 (Fig. 5.8.g-h) the differences tend to be of the order of magnitude of the *in situ* uncertainties, at some altitudes slightly smaller, and some slightly larger. The observed profile mean differences range between 24 and 52%.

Figure 5.9 presents the profiles of effective particle radius. We see that the (OE) satellite results are generally larger than the *in situ* radii (which is indicated by the relative differ-

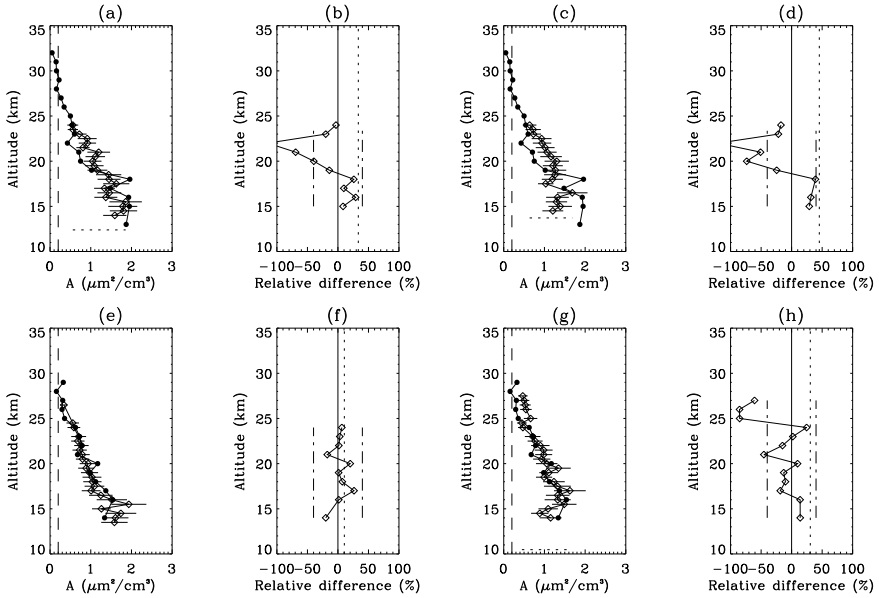

Figure 5.7: As Figure 5.5, but for surface area density as derived from the retrieved size distribution parameters. The *in situ* uncertainties (as marked by the dash-dotted vertical lines) are 40% [Deshler *et al., 2003*].

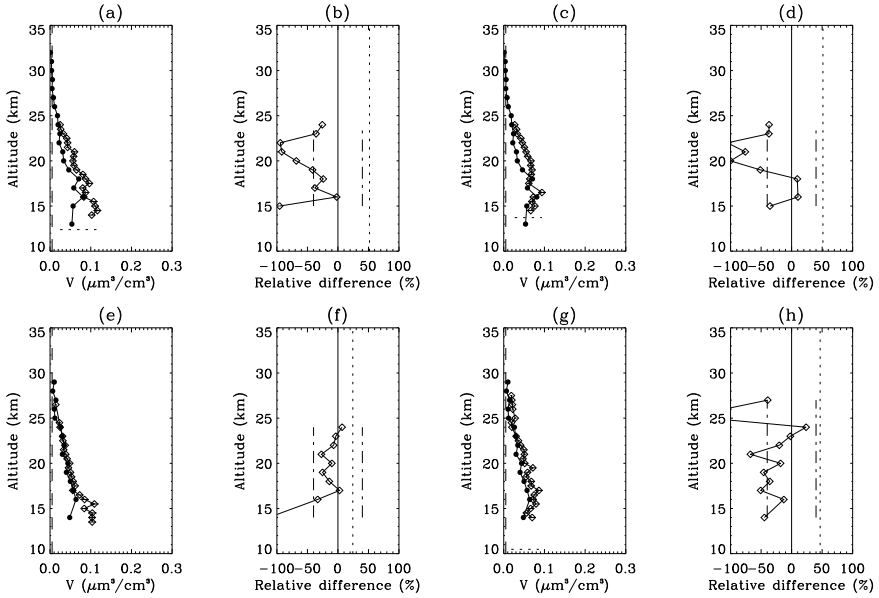

Figure 5.8: As Figure 5.5, but for volume density as derived from the retrieved size distribution parameters. The *in situ* uncertainties (as marked by the dash-dotted vertical lines) are 40% [Deshler *et al., 2003*].

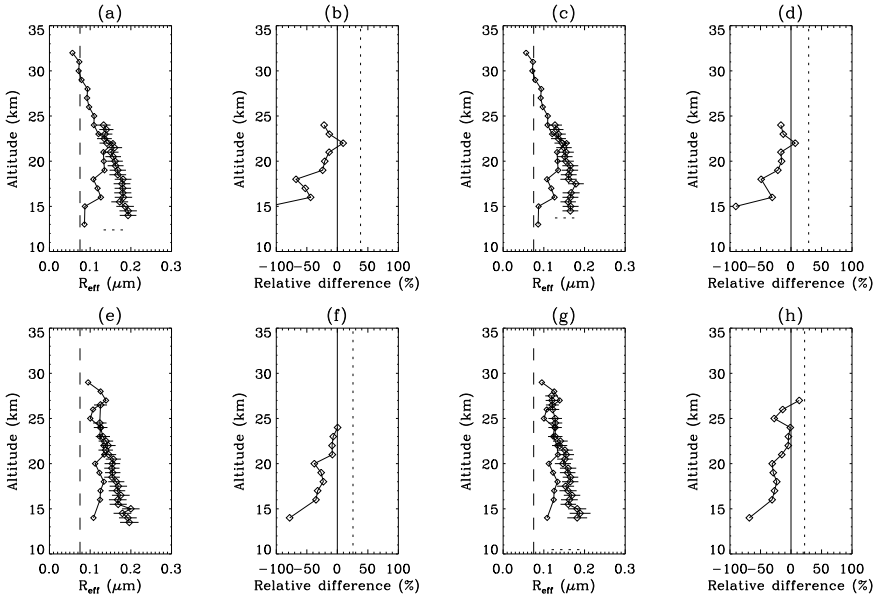

Figure 5.9: As Figure 5.5, but for effective radius as derived from the retrieved size distribution parameters.

ence being mostly negative) and closer to the *a priori* mean R_{eff} than the *in situ* data. In June (Fig. 5.9.a-d), the two profiles are observed to agree within the retrieved errors above 20 km, in December (Fig. 5.9.e-h) above 21 km. The profiles differ on average by between 22 and 38%.

Summary

The comparison between vertical profiles of aerosol properties as retrieved from SAGE II measurements (using the new Optimal Estimation retrieval algorithm) and coincident *in situ* data (as measured by Optical Particle Counters) has shown:

- The retrieved and the *in situ* profiles of aerosol properties are observed to be similar in their orders of magnitude and agreement between profiles within the retrieved and/or the *in situ* uncertainties is widely observed.
- The (absolute) differences tend to increase with decreasing altitude and are often largest at the lowermost heights, where the retrieved solutions tend to be closer to the *a priori* mean than the *in situ* values.
- The largest differences are observed where the *in situ* data are larger (smaller) than the *a priori* mean plus (minus) one *a priori* standard deviation. Such large and small values are unlikely to be retrieved due to the influence of the *a priori* constraint.

- In N, R, and A the differences tend to be equally positive and negative, meaning that no systematic bias can be observed, whereas the retrieved volume densities and effective radii are often larger than the *in situ* values.

- The profile mean differences range between 20-36% in number density, 9-13% in median radius, 10-46% in surface area density, 24-52% in volume density, and 22-38% in effective radius. Compared to the maxNS validation profiles (Sec. 4.5) these differences are slightly larger in N (12-34%, maxNS), R (4-7%, maxNS), and V (10-51%), smaller in A (11-71%, maxNS), and about twice as large in R_{eff} (4-15%, maxNS).

Overall, this profile comparison indicates that, despite the differences in measurement technique (and measurement volume), recording time and location, and despite the observed differences, the retrieved and the *in situ* profiles are similar. Assuming that the *in situ* data draw a realistic picture of the true distribution of aerosol properties, the results of this comparison indicate that the retrieved aerosol properties are also sensible. This means that the retrieved results have clearly improved on the *a priori* knowledge about the values and spatial distribution of stratospheric aerosol properties in June and December 1999. In Section 5.3 the above comparison with *in situ* measurements will also help to judge, whether or not the Optimal Estimation retrieval algorithm produces 'better' results than the Principal Component Analysis approach.

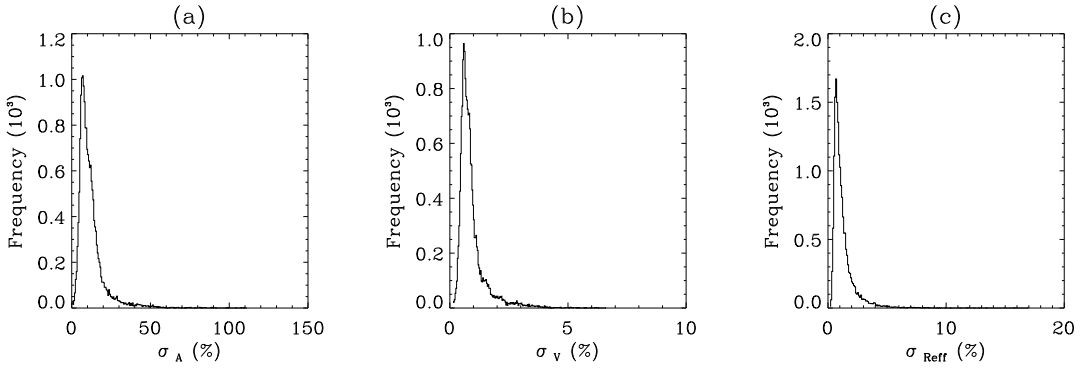

Figure 5.10: SAGE II, 12/1999: Relative uncertainties as estimated by the NASA LaRC (using PCA) of surface area density (A) and effective particle radius (R_{eff}). The associated uncertainties (standard deviations) in volume density (V) were derived from σ_A and σ_{Reff} according to Equation 5.4.

5.3 External Validation with PCA Retrieval Results

Since October 2001, the NASA Langley Research Center (LaRC) has made two aerosol retrieval products available to the scientific community, in addition to the already available spectral extinction data. These are aerosol surface area density and effective radius. These retrieval products were derived using Principal Component Analysis (PCA, see Sec. 2.4). From A and R_{eff} the associated volume densities can be derived using Eq. 2.11. The variance of $V = V(A, R_{\text{eff}})$ is given by

$$\sigma_V^2 = (\frac{\delta V}{\delta A})^2\sigma_A^2 + (\frac{\delta V}{\delta R_{\text{eff}}})^2\sigma_{\text{Reff}}^2 + 2\frac{\delta V}{\delta A}\frac{\delta V}{\delta R_{\text{eff}}}\sigma_{[\text{A,Reff}]}, \tag{5.3}$$

where $\sigma_{[\text{A,Reff}]}$ is the covariance of A and R_{eff}. This equation (see Taylor, 1939, Eq. 9.9) gives the standard deviation σ_V, whether or not the measurements of A and R_{eff} are independent or normally distributed. With $V = 1/3 \cdot A \cdot R_{\text{eff}}$ the associated variance in volume density is

$$\sigma_{\text{V}}^2 = \frac{1}{9}(R_{\text{eff}}^2\sigma_{\text{A}}^2 + A^2\sigma_{\text{Reff}}^2 + 2AR_{\text{eff}}\sigma_{[\text{A,Reff}]}), \tag{5.4}$$

and $\sigma_{[\text{A,Reff}]}$ is very small.

Figure 5.10 presents the uncertainties (due to the propagation of random error in the measured data) σ_{A} and σ_{Reff} as estimated by the NASA LaRC and the associated errors in volume density (derived using Eq. 5.4). In order to be directly comparable to the OE uncertainty estimates, an additional method bias error of up to a factor 2 (Thomason and Burton, 2006) would have to be added to the PCA uncertainties. It can be observed that the PCA uncertainties in surface area density are typically between 5 and 20%, with an

ensemble mean σ_A of 12%. The effective radius errors are observed to be smaller than 5% with an ensemble mean σ_{Reff} as low as 1.2%, and the derived volume errors are also very small with an ensemble mean σ_V less than 1%. These values can be compared to the OE retrieval noise S_m (Sec. 4.4.2), which is the retrieval uncertainty component due to measurement noise, which according to Section 4.4.4 is expected to be between 5% (maxNS) and 30% (minNS). In terms of the mean retrieved errors, this means comparing an error of 12% in A (PCA) with 1.15-6.9% (OE), comparing 1% in V (PCA) with 0.6-3.6% (OE), and comparing 1.2% in R_{eff} (PCA) with 0.65-3.9% (OE). While in A the PCA uncertainties are up to an order of magnitude larger than the OE uncertainty estimates, the random errors in V and R_{eff} are comparable in both methods.

PCA and *in situ* Vertical Profiles

The following comparison of profiles is performed in analogy to Section 5.2, where Optimal Estimation profiles were compared to correlative *in situ* data. The *in situ* profiles used in the comparison below are identical with the *in situ* profiles used in Figures 5.7-5.9 (Sec. 5.2), so that a cross-comparison can be performed to see whether the PCA or the OE method results tend to be closer to the *in situ* data.

Figure 5.11 presents the profiles of surface area density. Generally, the PCA surface area densities tend to be smaller than the OPC *in situ* values, with increasing deviations at lower altitudes. The PCA error bars are very small so that in this respect the PCA and the *in situ* profiles agree only at a few altitudes. With respect to the larger *in situ* uncertainties (Fig.s 5.11.b/d/f/h) more matches can be observed. Below approximately 19 km, the PCA surface areas are observed to be about 40-50% smaller than the *in situ* values. Similar biases have been reported by other researchers [Steele *et al.*, 1999; Deshler *et al.*, 2003]. Above 19 km, the differences between the PCA and the *in situ* data tend to be smaller. The profile mean differences are observed to range between 30 and 40%.
Comparison between the above PCA and *in situ* surface area densities and the *in situ* profiles of number density (Fig. 5.5) and particle median radius (Fig. 5.6) shows that the difference between the PCA and the *in situ* surface areas tends to be large where the *in situ* number density is large and the *in situ* particle size particularly small.
Cross-comparison with the Optimal Estimation profiles (Fig. 5.7) shows that the PCA and the OE retrieved surface area densities are similar in their vertical structure, although the PCA surface densities are generally smaller than the OE values. It can also be observed that the OE surface areas tend to match the *in situ* values better, except above 19 km in June (Fig. 5.11.a-d) where the PCA values of A are closer to the *in situ* data than the OE solutions.

Figure 5.12 shows the profiles of volume density as derived from the operationally retrieved surface area density and effective radius. In general, the PCA profiles of volume density

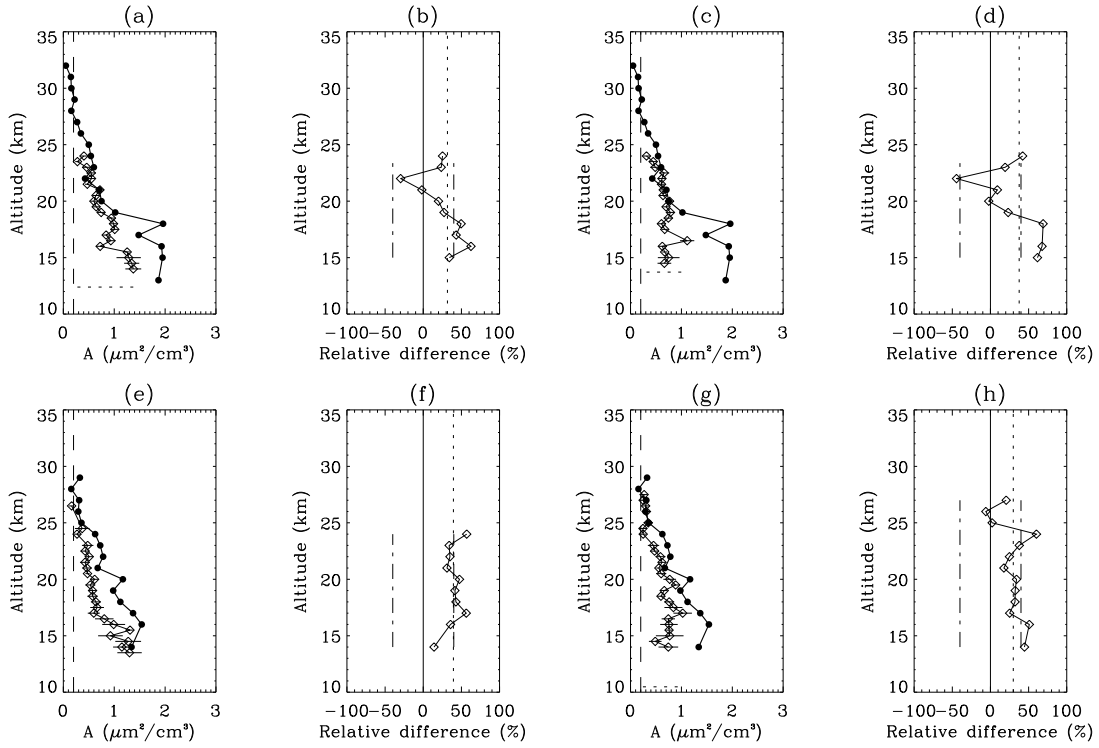

Figure 5.11: Vertical profiles of surface area density. Frames a/c/e/g: Surface areas as retrieved by the NASA LaRC using the PCA approach (diamonds with error bars, measured on June 22 (a) and June 23 (c), 1999) and correlative *in situ* measurements (without error bars, measured on June 23, 1999) of N; the vertical dashed line marks the *a priori* mean, and the short horizontal dotted line marks the tropopause level (NMC data) at the time of the SAGE II measurements. Frames b/d/f/h: Associated relative differences. The long vertical lines mark the zero (solid) and the profile mean difference (dotted). The *a priori* uncertainty of 40% [Deshler *et al., 2003*] is marked by the dash-dotted line.

are observed to match the *in situ* data well. At most altitudes the retrieved and *in situ* volumes are observed to agree within the PCA errors as well as within the *in situ* uncertainties, and the profile mean differences range between 20 and 29%. Comparison with the OE profiles (Fig. 5.8) shows that, as in the case of surface area, the retrieved PCA and OE volume profiles are very similar in their vertical structure, and that the PCA volumes tend to be slightly smaller than the OE volumes. This means that in the case of our four example profiles the PCA volumes tend to match the *in situ* data better than the Optimal Estimation retrieved results. This can also be seen in the profile mean differences, which are 20-29% (PCA) compared to 24-52% (OE).

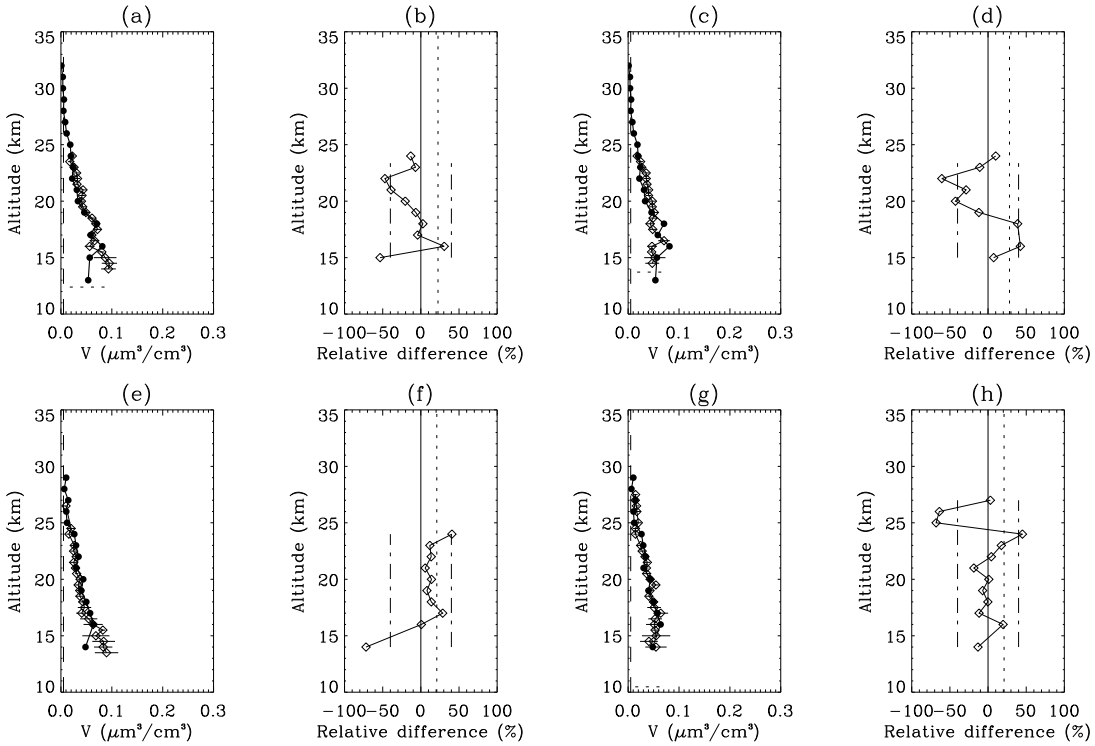

Figure 5.12: As Figure 5.11, but for volume density as derived from surface area density and effective radius. The *in situ* uncertainties (as marked by the dash-dotted vertical lines in frames b/d/f/h) are 40% [Deshler *et al., 2003*].

Figure 5.13 presents profiles of PCA effective particle radius and the correlative *in situ* profiles. It can be observed that the PCA effective radii tend to be nearly twice as large as the *in situ* values and the profiles do not agree anywhere. The average profile differences are 51-64%, which is about a factor two more than the average difference between the OE and the *in situ* profiles (Fig. 5.9) which are 22-38%.

Comparing OE and PCA Retrieval Results

In Figure 5.14 the PCA surface area densities, effective radii and the associated volume densities are presented in comparison with the Optimal Estimation solutions. It can be observed that the OE and the PCA results are correlated, but the OE surface areas and volumes tend to be generally larger and the effective radii generally smaller than the PCA values.

Figure 5.15 presents the relative differences (in %) between the respective method results,

$$\Delta x = \frac{(x_{\mathrm{OE}} - x_{\mathrm{PCA}})}{x_{\mathrm{OE}}} * 100. \tag{5.5}$$

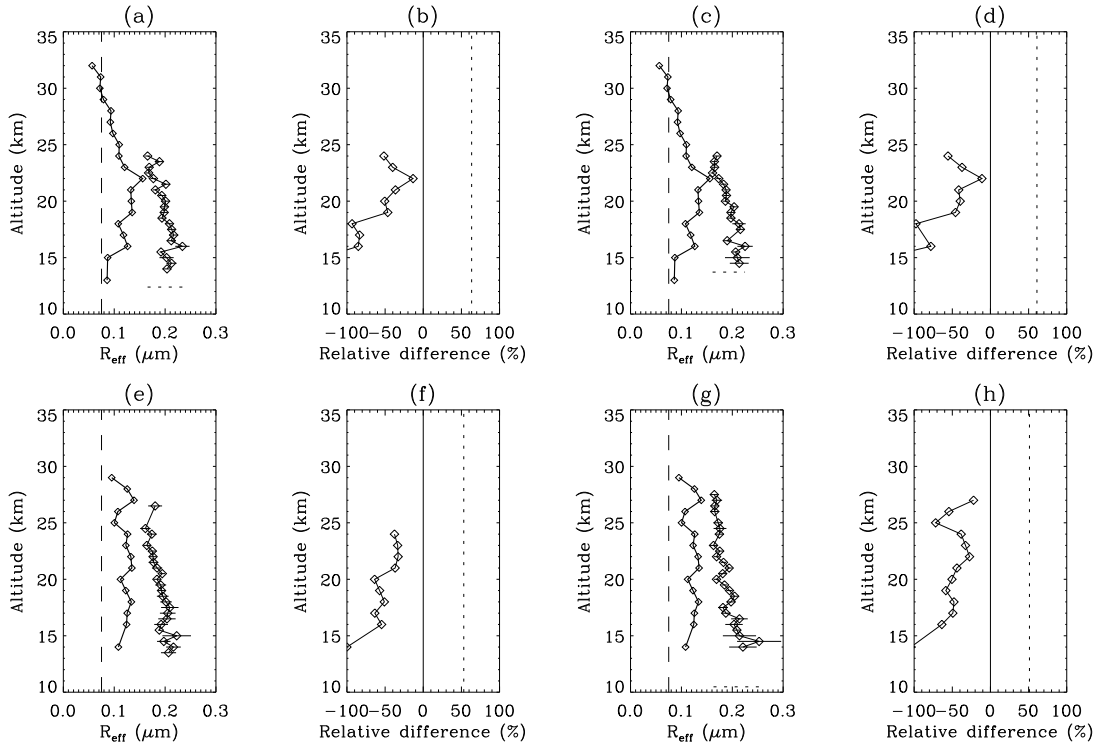

Figure 5.13: As Figure 5.11, but for effective particle radius.

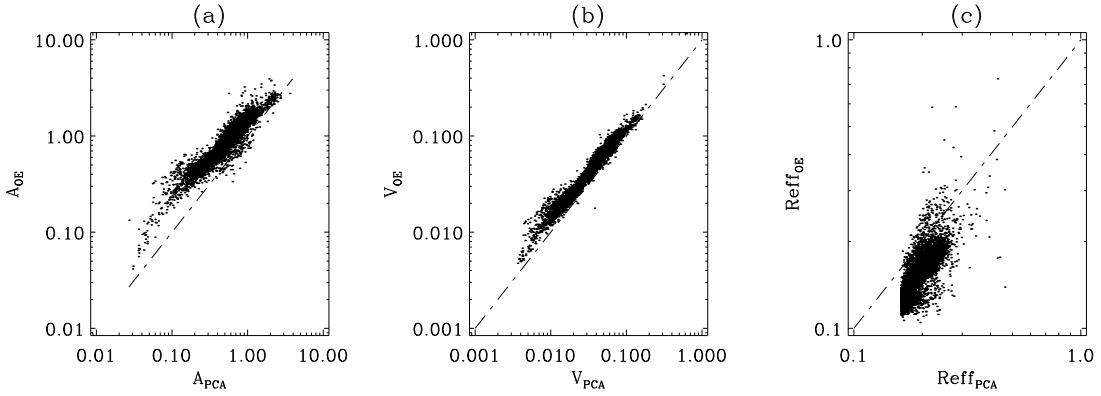

Figure 5.14: SAGE II measurements (December 1999): Principal Component Analysis (PCA) retrieval results of surface area density in $\mu m^2 cm^{-3}$, effective radius R_{eff} in μm (courtesy of NASA LaRC) and the associated volume density in $\mu m^3 cm^{-3}$, compared to the Optimal Estimation retrieval results. The diagonal line marks $x = y$ where both results would be identical.

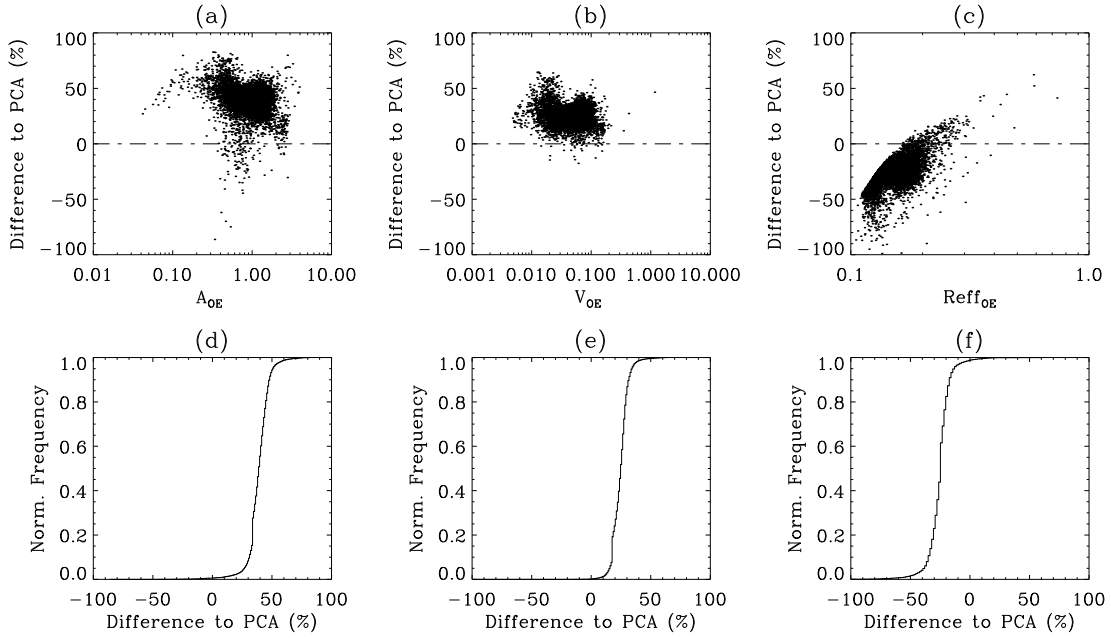

Figure 5.15: Difference (in %) – with respect to the Optimal Estimation (OE) results – between the retrieved A, V, and R_{eff} (OE, this work) and the Principal Component Analysis (PCA) results (courtesy of NASA LaRC). Frames a-c: Difference as a function of A, V, R_{eff}. Frames d-f: Cumulative histograms of the differences.

It can be observed that in effective radius (Fig. 5.15.c) the difference clearly increases toward small R_{eff}. In surface area density (Fig. 5.15.a) and volume density (Fig. 5.15.b) a similar tendency can be observed, even though it is not quite as strong as in effective radius. Figure 5.15.d shows that the great majority of OE surface areas are between 20 and 60% larger than the PCA surface areas. The OE volumes are observed to be mostly between 10 and 40% larger (Figure 5.15.e), and the OE effective radii tend to be 10 to 40% smaller (Figure 5.15.f) than the respective PCA values.

In Figure 5.16.d-f the difference between the two data sets is compared to the retrieved OE uncertainties in order to see whether some OE and PCA results agree within the estimated OE uncertainties (which are assumed to be a realistic or a slightly conservative estimate of the true errors). Ratios of true and retrieved uncertainties lying between -1 and $+1$ indicate that the OE and the PCA results agree within 2 standard deviations of the OE uncertainties. From the cumulative histograms it can be seen that in A approximately 80% of the data agree within two standard deviations, and in V and R_{eff} it is less than 50%. Since for a Gaussian distribution 95% of the data are expected to agree within two standard deviations, the smaller numbers indicate that the difference between the OE and the PCA integrated aerosol properties is significant.

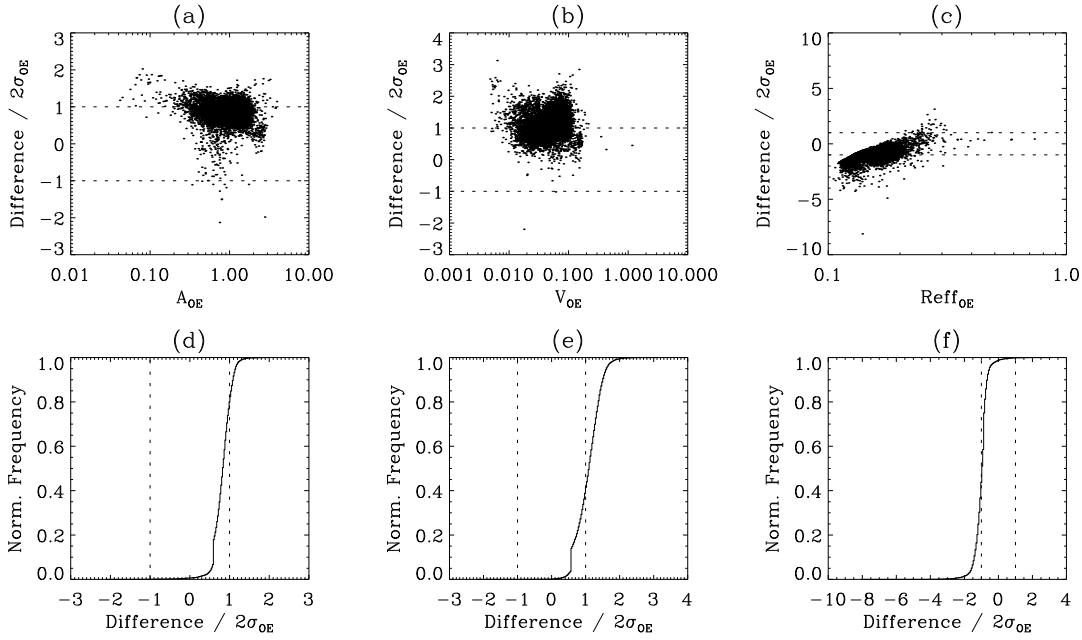

Figure 5.16: Difference (in %) – with respect to the Optimal Estimation (OE) results – between the retrieved A, V, and R_{eff} (OE, this work) and the Principal Component Analysis (PCA) results (courtesy of NASA LaRC) compared to the retrieved retrieved OE uncertainties (in %). The ratios are shown as a function of A, V, R_{eff}(frames a-c) and in the form of cumulative histograms (frames d-f).

Summary

When comparing four example vertical profiles of PCA and OE aerosol properties and correlative *in situ* data we observed that

- the PCA and OE profiles have similar vertical structures;
- the PCA volumes tend to match the *in situ* profiles well. They tend to be slightly smaller than the OE volumes which, in the case of our four example profiles, means that they match the *in situ* data slightly better than the OE retrieval results.
- the PCA effective radii differ strongly from the *in situ* profiles (on average 51-64% per profile), and by about twice as much as the OE profiles.
- the PCA surface area densities do not match the *in situ* data as well as the OE results, except in June above 19 km. They tend to be about half the size of the *in situ* values, which agrees with observations of similar biases reported by other researchers [Steele *et al.*, 1999; Deshler *et al.*, 2003].
- regions where the PCA surface areas are observed to strongly underestimate the *in*

situ surface areas tend to coincide with large *in situ* number densities of very small particles.

When directly comparing the PCA and the OE retrieved integrated aerosol properties we observed that the retrieved results are of the same order of magnitude but with the following differences:

- the OE surface area densities are typically larger (by 20-60%) than the PCA surface areas;
- the OE volume densities are typically larger (by 10-40%) than the PCA volumes;
- the OE effective radii are typically smaller (by 10-40%) than the PCA effective radii;
- the PCA error estimates represent the retrieval uncertainty due to the propagation of random error and should therefore be compared to the retrieval noise rather than the total retrieved uncertainty. While in A the PCA uncertainties are up to an order of magnitude larger than the OE uncertainty estimates (12% in PCA compared to 1.15-6.9% in OE), the random errors in V (1% in PCA compared to 0.6-3.6% in OE) and R_{eff} (1.2% in PCA compared to 0.65-3.9%) have similar magnitudes in both methods.

Thomason and Burton (2006) suggest that the PCA surface areas are smaller than the *in situ* measurements because of the low sensitivity of the SAGE II instrument to very small particles. The observations that in the presence of very small aerosol particles (a) the PCA aerosol properties tend to underestimate the correct surface area density, (b) the OE retrieval solutions of A are generally larger than the PCA values, and (c) the OE values tend to be closer to the *in situ* values than the PCA results, could therefore indicate that the Optimal Estimation algorithm provides a good way of filling 'the blind spot' with suitable information about the smallest particles.

5.4 Zonal Mean Aerosol Properties

In this section, the aerosol microphysical properties as retrieved from SAGE II measurements taken in the year 1999 are used to study seasonal differences on a global scale. For this purpose, latitude-altitude cross-sections of zonally averaged aerosol properties were generated for the months of March, June, September and December. The monthly means are calculated on a 10° latitude by 1 km altitude grid.

Latitudinal Sampling Characteristics

Whether or not those monthly means are statistically meaningful is determined by the variability of the averaged variables in combination with

1. the number of events averaged per grid box, which should be sufficiently large and ideally distributed evenly across the entire latitude-altitude grid, and with

2. the sampling dates (of all data in a particular bin), which ideally are representatively distributed across the entire averaging period.

These aspects will be looked at first in order to be able to draw meaningful conclusions from the monthly mean zonal cross-sections.

Figure 5.17 shows the geographical distribution of all locations associated with SAGE II measurements recorded in March, June, September, and December 1999. The geographical coverage as well as the data density at any particular location are observed to vary from month to month. In March and December, for instance, measurements exist as far south as 70°S, whereas in June 50°S is the southernmost latitude covered. It can also be observed that the latitudinal spacing increases toward small latitudes, and that at some latitudes the satellite crosses twice, whereas at other locations measurements are taken only once a month.

Figure 5.18 presents the number of measurements per grid box counted on a 10° by 1 km latitude-altitude grid. Naturally, the highest data densities are observed where the satellite crossed several times. This is generally around 50-70°S and 40-60°N, where counts are as high as 100 to 400. Counts as low as 10 measurements per bin are observed at the lower and upper ends of the measured profiles.

Figure 5.19 shows how the measurements are distributed throughout the averaging period of a month. The patterns are observed to differ from month to month. At some latitudes, the satellite measured several times during one month or during a prolonged period. At these locations the monthly means are likely to be more representative of the whole month than at other latitudes, where measurements are only available during a few consecutive days. At those locations the temporal variability of a variable determines whether or not a monthly mean value is representative of the entire month. If there is a noticeable

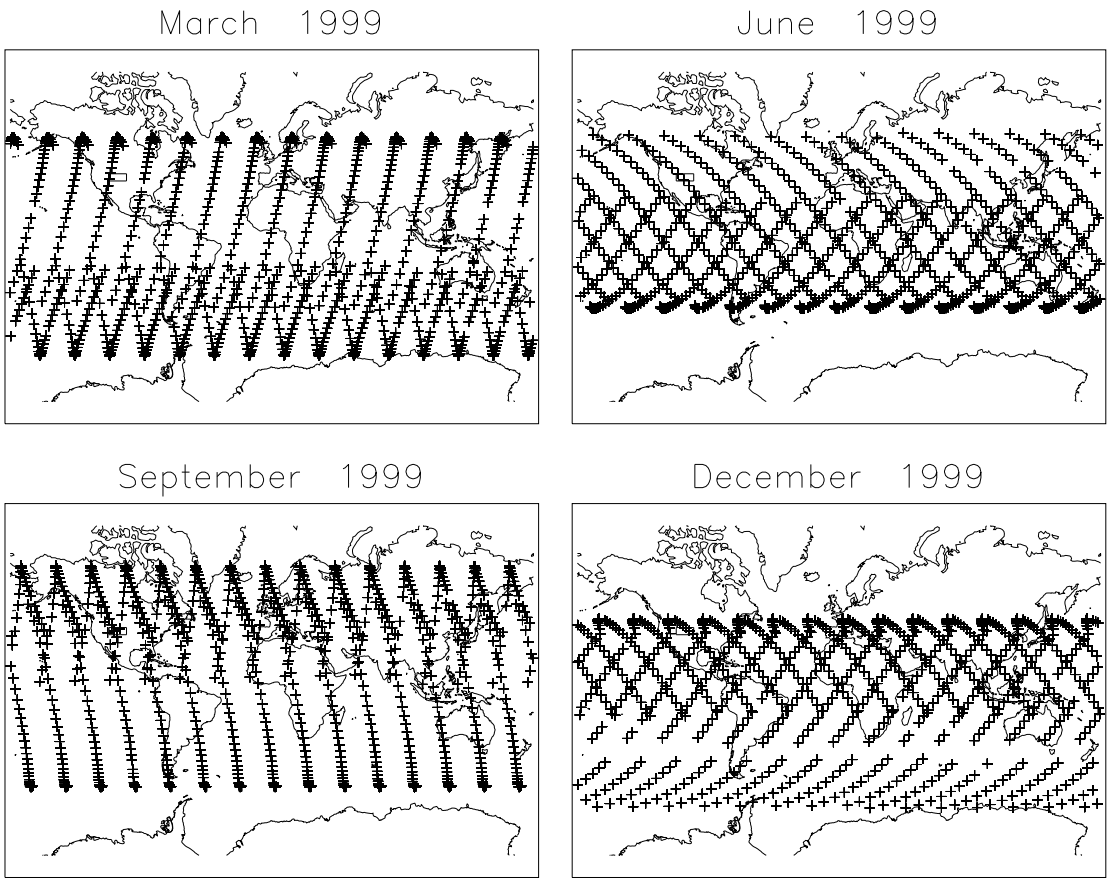

Figure 5.17: Locations of SAGE II measurements in March, June, September, and December 1999. The square marks the approximate location of Laramie/Wyoming (41°N, 105°W).

change with time, then the monthly mean may not be a representative value because we do not know how the variable changes during those other times (during the same averaging period) where no measurements exist. If the monthly means were not representative, then the temporally uneven sampling of the SAGE II satellite instrument could also lead to a *spatio-temporal bias* in the monthly latitude-altitude cross-sections because of the systematic temporal difference between the measurements taken at adjacent latitudes.
Considering both the data density (on the 10° latitude by 1 km altitude grid) and the spatial/temporal sampling aspect, we assume that the monthly means presented below are likely to be most meaningful at the following locations:

- in March: 70°S (14-23 km) to 30°S (18-27 km), and 50°N (13-23 km);
- in June: 50°S (14-26 km), 40°S (16-26 km), and 10 to 20°N (17-25 km);

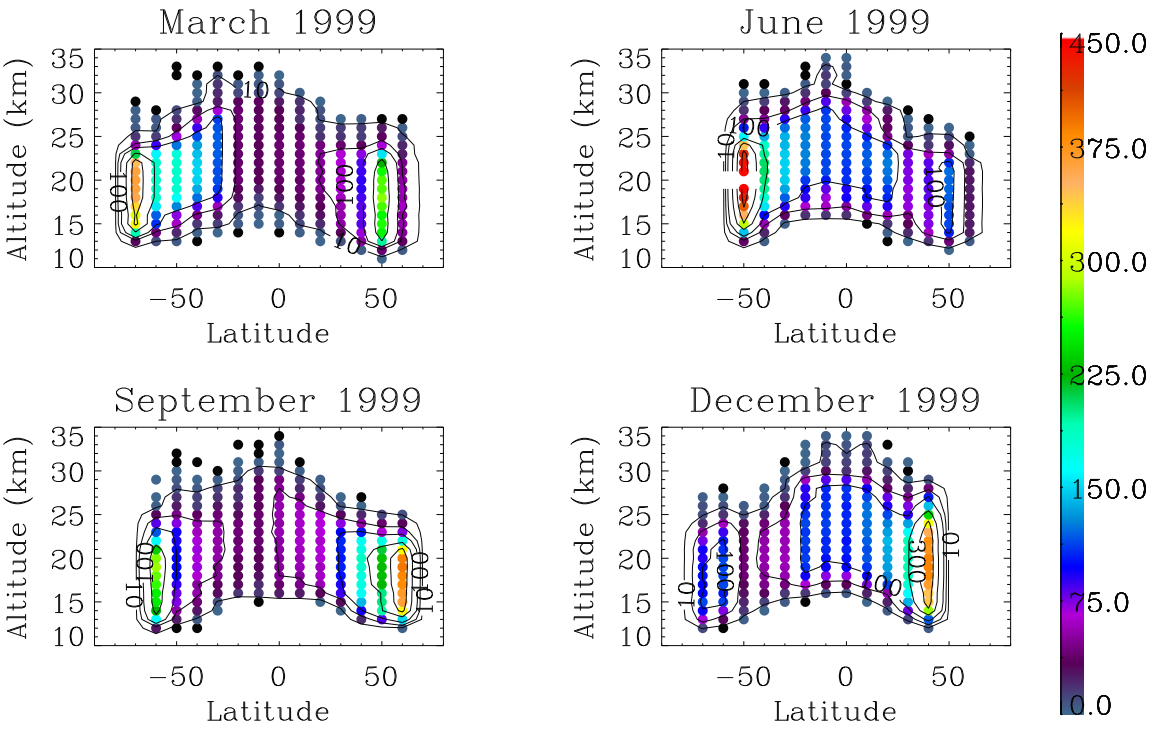

Figure 5.18: Number of data points per grid box for monthly zonal averages in the lower stratosphere. Each grid box is represented by a spot. Contours are marked for 10, 50, 100, 200, and 300 counts.

- in September: 30°N (15-23 km) to 60°N (13-23 km), and 60°S (13-24 km);
- in December: 30°N (14-25 km) to 40°N (13-27 km), and 20°S (18-26 km).

1999 Monthly Means

Before analysing the retrieved aerosol properties we start by looking at two input parameters the climatological distributions of which are generally better known than the aerosol properties: atmospheric temperature, and sulphuric acid concentration.

Figure 5.20 presents the zonally averaged distribution of temperature (from the National Meteorological Center, NMC) in the lower stratosphere. Generally, temperature is observed to be lowest in the tropics between 15 and 18 km and increases with increasing altitude and toward higher latitudes. A comparison between the June and the December monthly means shows that, as expected, the stratosphere is colder in the respective winter hemispheres. The temperature minimum observed south of 50°S in September is probably the edge of the Antarctic polar vortex. The general structure of the zonal

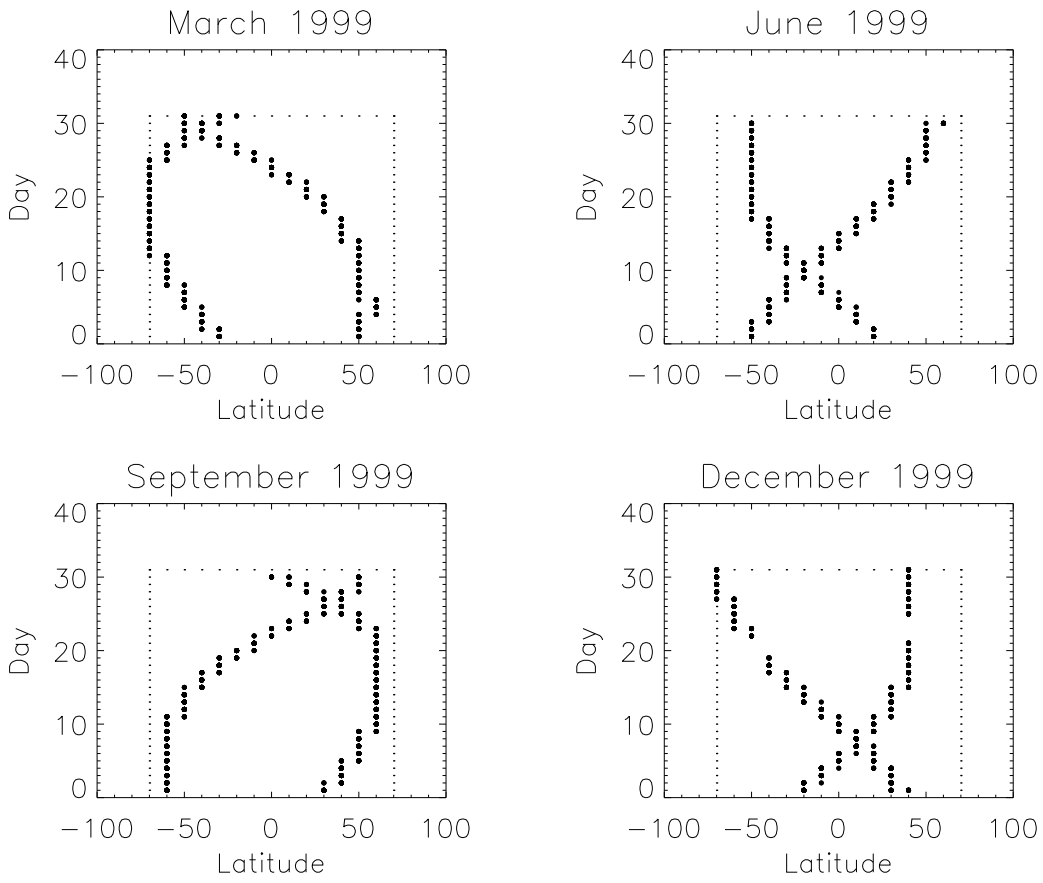

Figure 5.19: Monthly measurement dates as a function of latitude in March, June, September, and December 1999 for a 10° by 1 km latitude-altitude grid, resulting from the SAGE II sampling characteristics.

mean temperature distribution is very similar to the observed monthly and zonally averaged temperature distribution presented by Holton *et al.* (1995, Fig.12.2), which can be explained by a balance between infrared radiative cooling and radiative heating owing to the absorption of solar ultraviolet radiation by ozone.

Figure 5.21 shows monthly mean sulphuric acid concentrations (H_2SO_4 in weight percent) that were calculated as described in Sec. 2.2 from the observed pressure and humidity (SAGE II) and temperature data (NMC). The mean acidities are observed to closely follow the temperature distribution (Fig. 5.20) with low values in the lower tropical stratosphere (55%) and a general increase with altitude (more than 80% above 30 km). Comparison with Figure 5.20 shows that low temperatures are generally associated with low acid weight percentages. Seasonal variations in average droplet acidity are mainly observed poleward of 50° N and S, where low values occur in the winter hemisphere, and high values in the

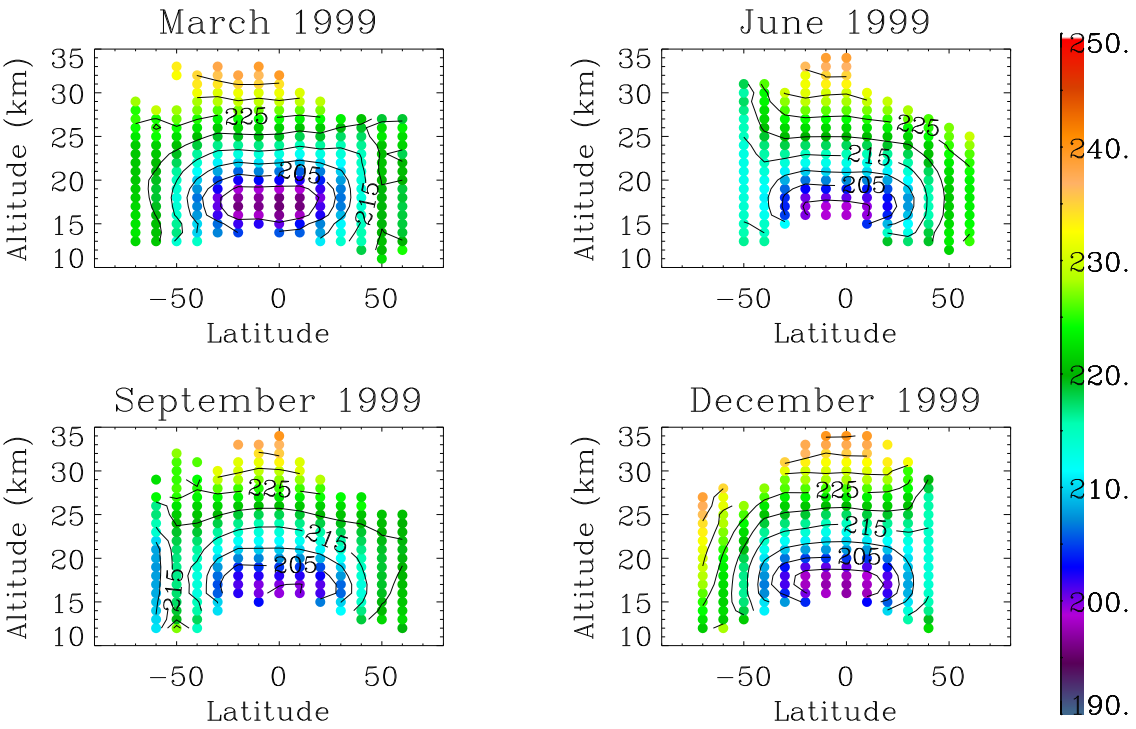

Figure 5.20: Lower stratosphere zonal mean temperature T in K (NMC data).

respective summer hemisphere. These monthly means are very similar to the five year (1986-1990) zonal mean averages presented by Yue *et al.* (1994, Fig. 2a).
The similarity with values in the literature in both temperature and sulphuric acid concentration is encouraging and indicates that the monthly averages are representative despite the uneven spatial and temporal distribution of the SAGE II measurements from which the above monthly means were generated. We therefore move on to evaluate the monthly mean retrieved aerosol properties.

Zonal Mean Aerosol Properties

For the presentation of the monthly mean retrieved aerosol properties a linear interpolation scheme was used to generate continuously coloured contour plots. The scheme works well except at the edges of the data grid, where extrapolation beyond the original grid tends to produce unrealistic values. Some values near the upper and lower end of the profiles are also less meaningful because of a low number of measurements averaged or high uncertainties associated with the mean. Therefore, a mask was applied to the Figures 5.22-5.22 in order to block out these less meaningful data points.

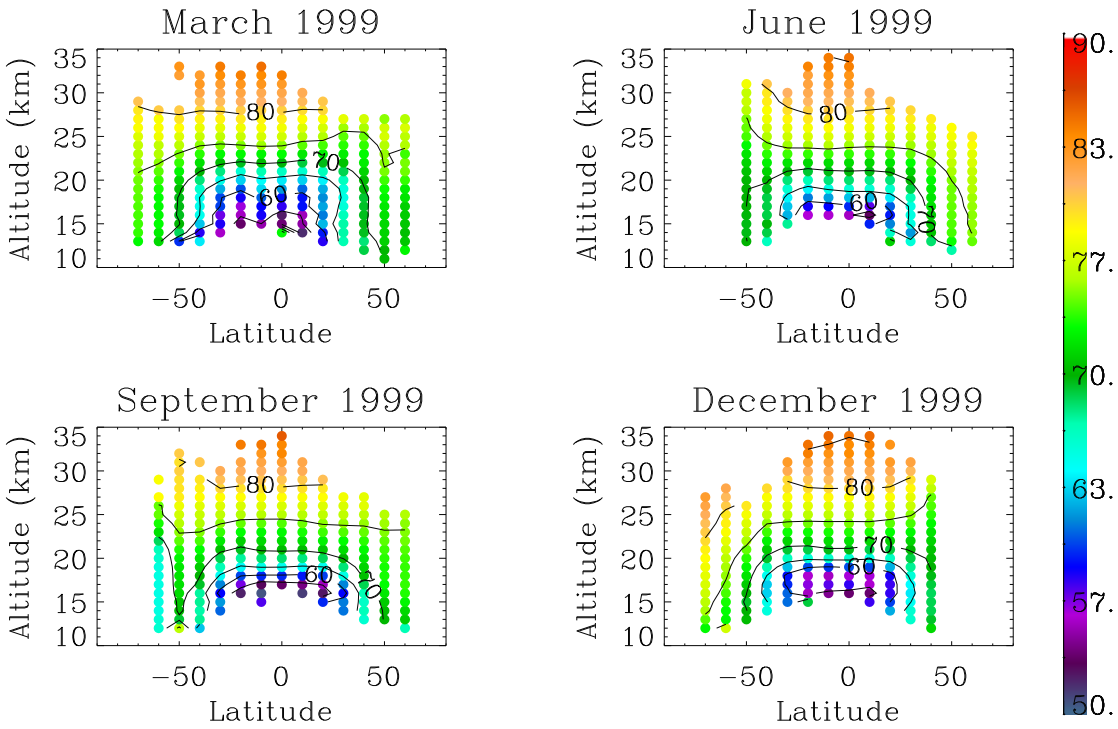

Figure 5.21: Zonal mean H_2SO_4 concentration (in % by weight of H_2SO_4 in H_2SO_4-H_2O aerosols) in the lower stratosphere.

Median Particle Radius

Figure 5.22 presents monthly means of the retrieved and zonally averaged stratospheric median particle radius. Generally, the retrieved particle radii are observed to range between approximately 0.07-0.09 μm near the tropopause and 0.05 μm near 30 km. The highest values are observed at about 40-60°N/S near the tropopause. Small particles are observed at the upper ends of the grid, and between approximately 18 and 23 km in the tropics. The latter is a local minimum, where at an altitude of 20 km the particles can be nearly half the size of those particles observed at high latitudes.

The largest seasonal differences are observed in the mid-latitude lower stratosphere below 23 km, where particle size maxima are observed in late winter and early summer, that is in September and December in the southern hemisphere (SH), and in March and June in the northern hemisphere (NH). Mid-latitude seasonal particle size minima are observed in late summer and early winter, that is in March and June in the SH, and September and December in the NH. The spring (maximum) to autumn (minimum) seasonal differences at mid-latitudes (for instance near 50°N/S at 20 km) tend to be more prominent in the southern hemisphere than in the northern hemisphere.

A tropical particle size minimum (of 0.04-0.05 μm near 20 km) is observed in all four

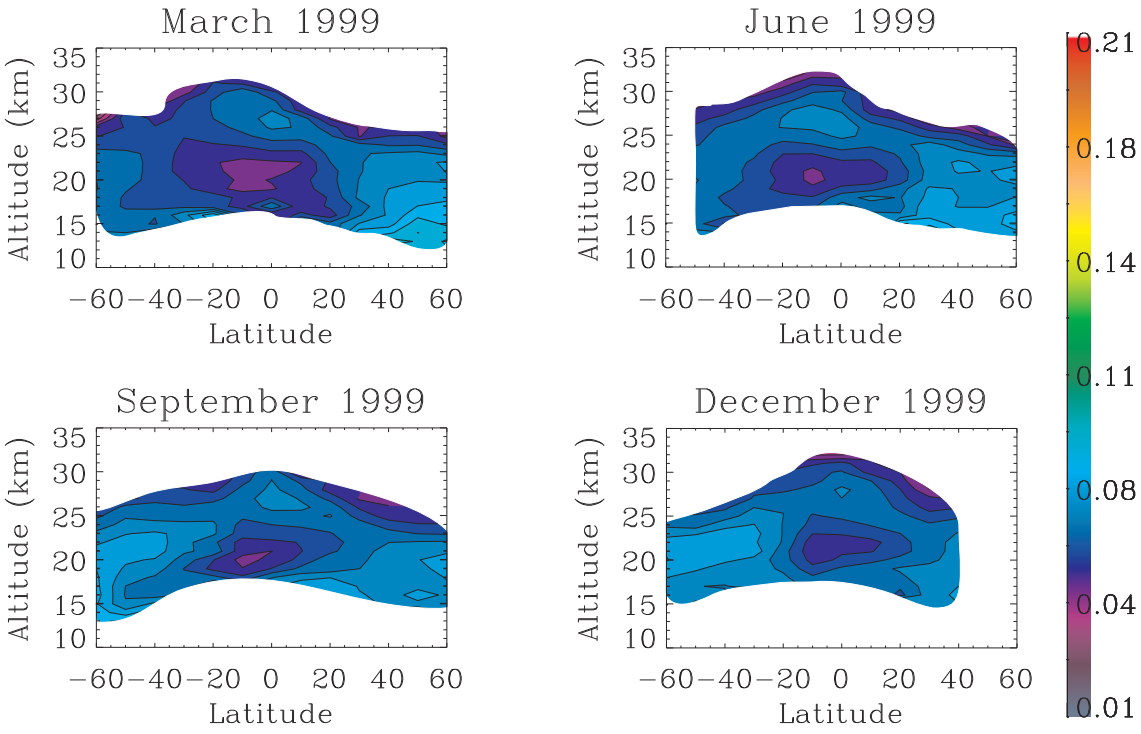

Figure 5.22: Zonal mean retrieved stratospheric median radius R in μm.

seasons even though with variable depth and asymmetrical extent downward into the subtropical lower stratosphere. In December and March the particle size minimum extends into the northern lower subtropical stratosphere, whereas in June and September the smaller particles are observed in the southern hemisphere lower subtropical stratosphere. In the northern hemisphere this causes a subtropical winter minimum.

Comparison with aerosol composition (Fig. 5.21) and atmospheric temperature (Fig. 5.20) shows that the seasonal differences in the particle size are more pronounced than those observed in the other two variables. Near 60°S at 20 km, for instance, the radius grows from 0.07 μm in March to 0.09 μm in September, which corresponds to a 25% change (with respect to the arithmetic mean), whereas at the same time the composition drops from about 73 to 68%, which corresponds to less than 15% change, and the temperature drops from 220 to 210 K which is less than 5%. The faster a variable changes, the more it is prone to spatio-temporal averaging biases on the latitude-altitude grid that may arise as a consequence of the measurements being not necessarily representative of the entire averaging period. Potentially affected in this respect are therefore the gradient structures of the mid-latitude winter/spring maxima in March and September, because there is a systematic difference between each measurement location and the associated recording time (Fig. 5.19). Since in those areas measurements exist only during a few consecutive

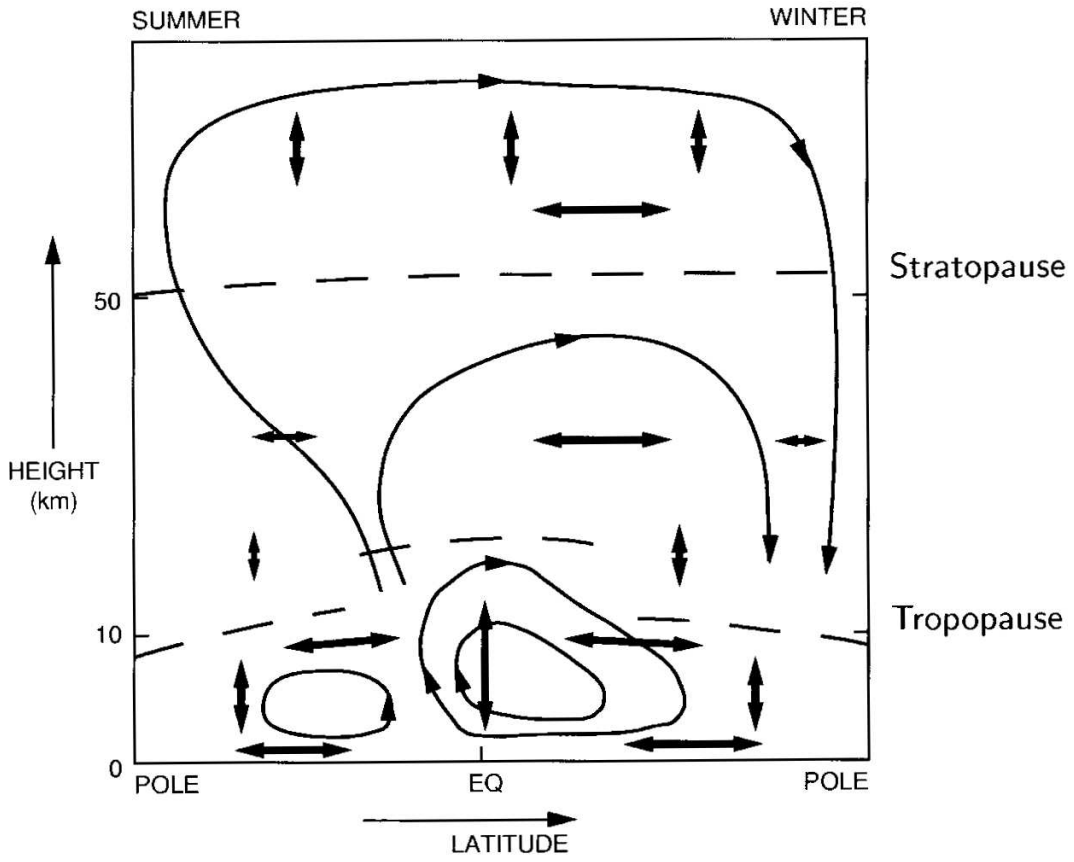

Figure 5.23: Schematic picture of the meridional circulation in the atmosphere (WMO 1995, as shown in Labitzke and van Loon, 1999).

days and nothing is known about the variability during the rest of the month, the data at hand do not allow us to quantify such a potential spatio-temporal bias. A bias of this kind can be eliminated by using measurements that are more evenly spaced in both time and space. In the case of SAGE II this can be achieved by generating climatological means using several years worth of data. However, this would also mean smoothing out the characteristics of individual years.

Interpretation

An increase in the median particle size can be caused by the growth of existing particles or by advection of large particles. In the same way the median particle size can decrease due to particles shrinking through the loss of material by evaporation, due to the advection of many small particles and/or due to new particle formation. From the point of view of atmospheric dynamics, two processes are likely to influence the distribution of aerosol

properties: the supply of material by advective transport within the stratosphere, and dynamic mixing between the upper troposphere and the lower stratosphere. A schematic of the average meridional circulation in the stratosphere, also known as the *Brewer-Dobson Circulation*, is presented in Figure 5.23. This large-scale circulation is characterised by rising air in the subtropics of the summer hemisphere and poleward and downward transport into the winter hemisphere. This generally poleward mass transport can be inhibited in zones where the potential vorticity is enhanced. A strong barrier exists for instance near 60° latitude in winter (above all in the southern hemisphere), where the polar vortex effectively obstructs transport to higher latitudes. As a consequence, a local subsidence of air and other constituents is induced just above the tropopause, leading to a localised accumulation of material in the lower stratosphere.

In a volcanically perturbed atmosphere, where aerosol rich air and large particles are transported from the equatorial lower stratosphere ('tropical reservoir') down and to higher latitudes, a winter maximum in all aerosol properties can be explained this way: the number of particles increase due to the accumulation of particles [Bingen *et al., 2004a*], the median radius increases due to the advected larger particles, and consequently aerosol surface area and volume will increase as well. Under low aerosol loading (background) conditions, however, subsidence in the winter hemisphere (at mid- to high latitudes) means bringing smaller particles from above to lower altitudes, which would tend to decrease the median particle radius. The observation of a late winter maximum (in Figure 5.22) therefore indicates that processes other than advection determine the particle size development.

Comparison with aerosol composition (Fig. 5.21) and atmospheric temperature (Fig. 5.20) indicates that, in general, small particles tend to be associated with warmer temperatures and higher acidities, whereas larger particles are observed at lower temperatures and lower acidities. This means that the particles could have grown by condensation of water vapour onto existing particles. In winter (summer) the particles are likely to grow (shrink) through adsorption (evaporation) of water molecules as the temperature and weight fraction of H_2SO_4 in the sulphuric acid (H_2SO_4-H_2O) binary aerosol decrease. If we assume that the amount of condensed sulphuric acid is approximately constant (Don Grainger, personal communication), then (knowing that H_2SO_4 has a higher molecular mass than water) the weight fraction of H_2SO_4 in H_2SO_4/H_2O can only decrease if many water molecules are adsorbed, which will increase the particle size.

In conclusion, the change at mid-latitudes from a southern early winter (June) lower stratospheric particle size minimum to a late winter (September) maximum suggests that in winter, particle growth processes dominate over the accumulation of smaller particles advected from above (which would decrease rather than increase the median particle size). In the summer hemisphere, which is generally characterized by upward motion, we would expect an increase in median particle size due to advection of larger particles from the tropopause region. However, the observed change at mid-latitudes from a southern early

summer (December) lower stratospheric particle size maximum to a late summer (March) minimum indicates that the associated decrease in particle median size is more likely the result of the particles shrinking due to evaporation of water vapour as the temperature increases. The observed simultaneous increase in acidity during summer suggests that the sulphuric acid component tends to be comparatively weakly variable (or possibly even conserved).

The observation that the tropical minimum (near 20 km) tends to be asymmetrical in its extent downward and into the subtropical lower stratosphere, is likely to be explained by the large scale circulation. The latter is directed northward in December through to March, and southward in June through to September, into the respective winter hemispheres, where there is a general wintertime subsidence that brings smaller particles from above down to lower altitudes.

Comparison with other Data Sets

Bingen *et al.* (2004a and 2004b) have generated a global climatology of aerosol size distribution parameters based on SAGE II measurements between 1984 and 2000. They applied a regularized inversion retrieval technique (see Sec. 2.4). With respect to seasonal differences, Bingen *et al.* (2004a) observed a clear seasonal cycle in the lower stratosphere (below 22 km) of both hemispheres. In the northern hemisphere mid-latitude lower stratosphere (at 40 to 50°N at 17.5 km) in 1999, R was observed to take its highest values in spring (April/May), and its lowest values in autumn (November/December), which agrees with our observations. In the southern hemisphere (at 40 to 50°S), Bingen *et al.* (2004a) observed the largest median particle radii in early autumn (March), and lowest R in winter (July/August). This, is not confirmed by our observations which indicate an autumn low (March/June) and a spring high (September/December). A theory as to how these changes can possibly be explained, is not offered by Bingen *et al.* (2004a). It would be interesting to know if their data show a similar cycle at other altitudes as well.

With respect to the observed particle sizes, comparison with the Optimal Estimation results shows that the median radii as retrieved by Bingen *et al.* (2004a) are about three times the Optimal Estimation results. For instance, in 1999, at an altitude of 17.5 km, the mid-latitude (40 to 70°N/S) median particle radii retrieved by Bingen *et al.* (2004a) range between 0.25 and 0.33 μm (NH), and between 0.27 and 0.37 μm (SH), whereas the OE results are on the order of 0.08 μm at the same locations. Simultaneously, the particle number densities retrieved by Bingen *et al.* (2004b) are observed to be low compared to *in situ* Optical Particle Counter measurements. The authors ascribe the difference to the difficulty of satellite instruments to measure the very small particles [Bingen *et al., 2004b*]. A comparison with the *in situ* monomodal and bimodal particle median radii shown in Figure 5.6 indicates that the retrieved OE values are more typical of monomodal background aerosols than those values presented by Bingen *et al.* (2004b).

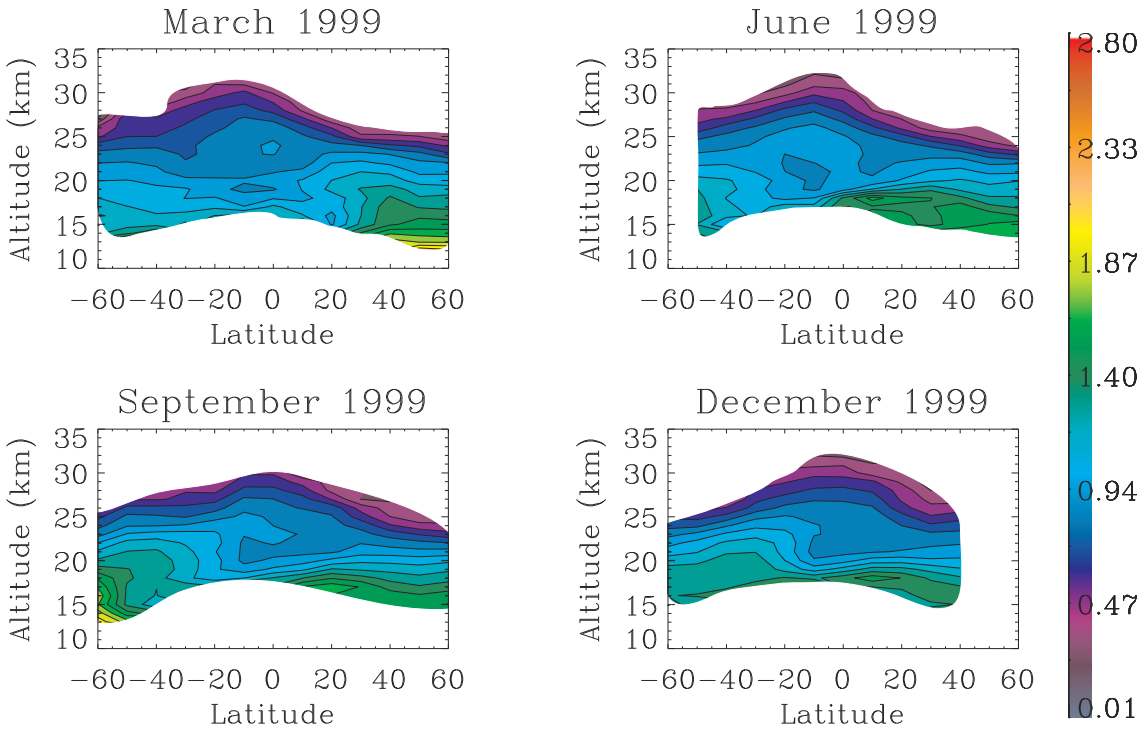

Figure 5.24: Zonal mean stratospheric surface area density A (in $\mu m^2 cm^{-3}$) in the lower stratosphere, as derived from the retrieved size distribution parameters. Colouring at the edge (where it continues beyond the contour lines) is adversely affected by rather unrealistic extrapolation results and should be disregarded.

Surface Area Density

Figure 5.24 displays monthly averages of surface area density as derived from the retrieved size distribution parameters. It can be observed that the spatial and temporal distributions of highs and lows are similar to those observed in median particle radius (Fig. 5.22). Above approximately 25 km, the contour lines follow the shape of the tropopause, which is higher in the tropics and lower at high latitudes. Consequently, aerosol surface area above 25 km density generally decreases toward higher latitudes. It can also be observed that the approximate centre of symmetry at this altitude is located slightly to the south of the equator, just like the intertropical convergence zone, which is the centre of symmetry of the global scale general atmospheric circulation. There is a general decrease with altitude, except in the tropics, where a local minimum is observed near 20 km. Values are observed to range between 0.4 μm^2 cm^{-3} near 30 km and 1.4-1.8 μm^2 cm^{-3} near the tropopause. At 20 km surface area ranges between 0.9 μm^2 cm^{-3} in the tropics and 1.3 μm^2 cm^{-3} near 60°S.

Seasonal differences are observed above all in the mid- to high-latitude lower stratosphere below approximately 20 km, where the month to month changes tend to be stronger in

the southern than in the northern hemisphere. Poleward of 40°S and N, large surface areas tend to occur in the respective late winter/early spring hemispheres of March and September. A similar observation has also been reported by Thomason *et al.* (1997). Their work analysed time series of surface area density (estimated from SAGE II and *in situ* data measured near Laramie/Wyoming, 41°N) at several heights (15, 20, 25 km) for the first few years after the June 1991 Pinatubo eruption, and observed a seasonal cycle with a wintertime maximum and a summertime minimum. Thomason *et al.* (1997) suggest that this seasonal cycle, observed in the post-volcanic data set, may be the product of aerosol transport from the tropics, which is most efficient in winter, combined with the replenishment of the lower stratosphere at middle and high latitudes by diabatic subsidence occurring in the winter hemisphere. This explanation sounds reasonable considering that in the aftermath of the Mount Pinatubo eruption the largest particles were observed in the tropical reservoir, and hence large particles are transported to higher latitudes through the large scale circulation. As in contrast, large scale transport and subsidence under low aerosol loading conditions means advecting smaller particles to higher latitudes, the seasonal maxima as observed in Figure 5.24 are more likely to arise from a combination of accumulated material at the dynamical vortex barrier and particle growth processes.
At 20°N near 17 km in March, a late winter/early spring minimum is observed at the time of the annual maximum observed near 40°N at 20 km. A similar minimum was already observed in median particle radius. Since the number of observations in that location is, however, rather small, this observation is taken note of but not further emphasized.

Volume Density

Figure 5.25 presents the lower stratosphere latitude-altitude cross-sections of zonally averaged volume density. As expected, the distribution of maxima and minima as well as the locations of pronounced seasonal variability closely follow the features observed in median particle radius and surface area density: seasonal difference are above all observed near the tropopause and at mid to high latitudes. Polewards of 40°N and S, those differences tend to be stronger in the southern than in the northern hemisphere. The observed volume densities range between 0.11 $\mu m^3 cm^{-3}$ near the tropopause and 0.01 $\mu m^3 cm^{-3}$ near 30 km.

Effective Radius

Figure 5.26 shows the associated zonal mean effective particle radius in March, June, September and December 1999. As all previously described aerosol properties, the effective particle sizes, too, are observed to generally decrease with altitude. Values range between approximately 0.2 μm near the tropopause to 0.1 μm near 30 km. This is approximately an order of magnitude larger than the median radius (Fig. 5.22). As to the annual

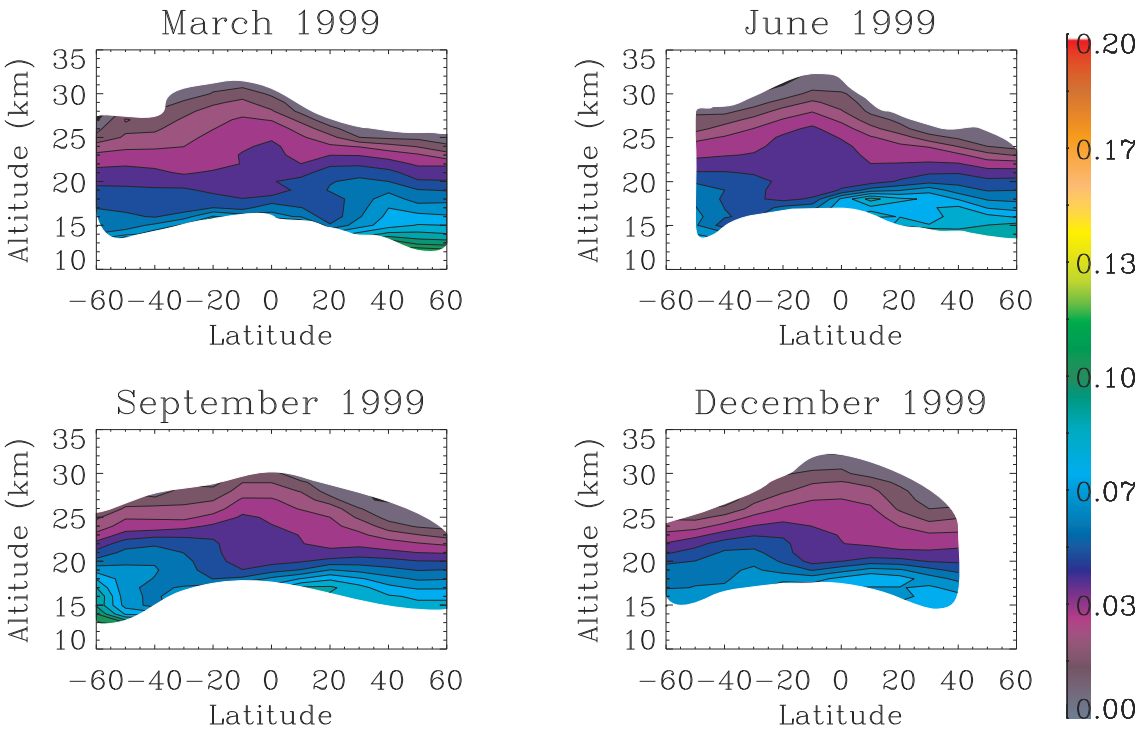

Figure 5.25: Zonal mean retrieved stratospheric volume density V (in $\mu m^3 cm^{-3}$) in the lower stratosphere.

cycle in R_{eff}, an approximate equatorial symmetry at mid-to high latitudes is retained throughout the year. Compared to surface area density or volume density little seasonal change is observed in the effective radius. This is a result of the seasonal differences in V and A being similarly strong, the ratio of which is proportional to the effective radius. The observation that the effective radius is clearly smaller than the theoretical threshold of 2 μm indicates (according to Lacis *et al.*, 1992) that under background conditions, stratospheric aerosol particles act to cool the stratosphere (Sec. 2.1). This is as expected, because small sulphuric acid particles are practically pure scatterers.

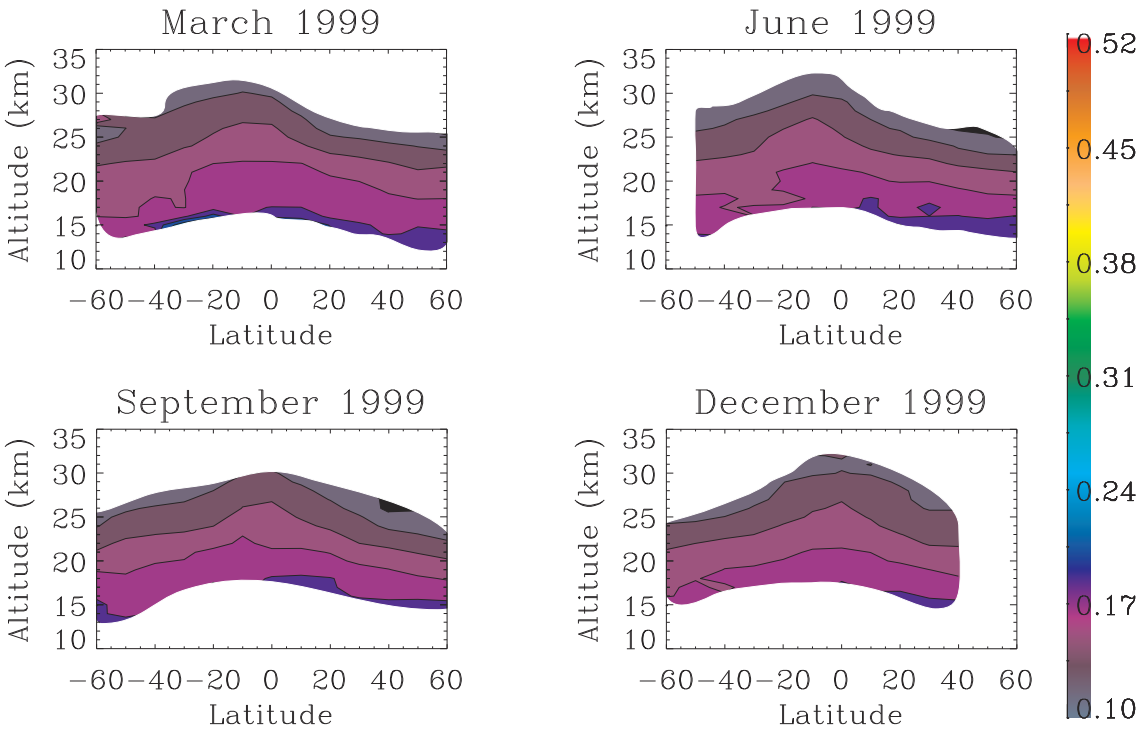

Figure 5.26: Zonal mean effective radius R_{eff} (in μm) in the lower stratosphere.

Summary

From the retrieved aerosol properties, monthly mean latitude-altitude cross sections were generated on a 10° by 1 km grid. Due to the SAGE II sampling characteristics, the measurements are unevenly distributed in both space and time, which results in areas with high data density, and other areas where the monthly average is based on only a few measurements. When analysing the monthly means and seasonal changes, more emphasis was put on observations made in areas of high data density. Our observations include the following main aspects:

- Low values in particle size and the derived integrated aerosol properties (surface area density, volume density, and effective radius) are generally observed at the uppermost altitudes of the available data (that is near 30 km).
- These are associated with comparatively high values in aerosol acidity and (ambient) temperature.
- The ambient temperature and aerosol acidity monthly mean cross sections agree with values in the literature with respect to both the distribution of the high and low values and the seasonal variability, which indicates that the monthly averages are representative despite the uneven spatial and temporal distribution of the SAGE II

measurements from which the monthly means were generated.

- The monthly means of aerosol particle median radius, surface area density, and volume density exhibit more variability and seasonal differences than the mean temperature and aerosol acidity. For instance, near 60°S at 20 km the median radius is observed to increase by 25% (with respect to the arithmetic mean) between March and September, whereas at the same time the composition and temperature drop by less than respectively 15% and 5%.

- The monthly amount of data available in 1999 does not allow us to quantify whether or not an aerosol property changes noticeably within a month. Consequently, a bias that could arise in connection with the temporally uneven sampling statistics cannot be excluded.

- Strong month-to-month differences are generally observed in the mid-latitude lower stratosphere below 23 km, where high values in aerosol median radius, surface area density, and volume density, are mainly observed in spring, whereas low values are recorded in autumn. Table 5.6 lists those observations month by month and for both hemispheres. In the northern subtropics, a different cycle is observed, with high values in (early) spring rather than in autumn, and with low values in (early) autumn rather than in spring. Those results are noted but not further emphasized because of the low data density in this area.

- The effective particle radius does not show much seasonal change.

These observations were compared to median radii as retrieved by Bingen *et al.* (2004a and 2004b) from SAGE II data between 1984 and 2000 and to surface area density as retrieved by Thomason *et al.* (1997) from SAGE II data measured during the first few years after the Mount Pinatubo eruption. We found:

- The median radii as retrieved by Bingen *et al.* (2004a) in 1999 are larger (by about a factor 3) than the Optimal Estimation results. Since their values were observed to overestimate *in situ* measurements, Bingen *et al.* (2004a) attribute the overestimated particle sizes to an underestimate in their number densities, which they explain by the low sensitivity of SAGE II data to very small particles.

- The annual cycle in median radius as observed by Bingen *et al.* (2004a) in the mid-latitude lower stratosphere (between 40 and 50°N at 17.5 km) in 1999 agrees with our observations, showing a spring maximum and an autumn minimum, whereas

- in the southern hemisphere our observations differ from their results, the latter indicating an early autumn maximum and a winter minimum (between 40 and 50°S at 17.5 km).

	March (early autumn)	June (late autumn)	September (early spring)	December (late spring)
20°S, 17 km:	low	low	high	high
40°S, 20 km:	low	low	high	high
60°S, 15 km:	low	low	high	high
	March (early spring)	June (late spring)	September (early autumn)	December (late autumn)
>40°N, ≤20 km:	high	high	low	low
20°N, 17 km:	low	medium	high	medium

Table 5.6: Seasonal differences in the retrieved zonal mean aerosol properties, as observed in the lower stratosphere in 1999.

- The annual cycle in surface area density as observed by Thomason *et al.* (1997) is characterized by a wintertime maximum and a summertime minimum, which is similar to the seasonal patterns observed in our Optimal Estimation monthly means.

The mechanisms that cause the observed seasonal differences can not be identified. Thomason *et al.* (1997) suggest that (at mid- and high latitudes) seasonal surface area density maxima (as observed in the aftermath of the large eruption of Mount Pinatubo) result from advective transport of large particles from the tropics. However, since away from any volcanic influence dynamic transport from the tropics means bringing smaller particles to higher latitudes, we suggest that the observed seasonal differences are more likely to be explained by particle growth/shrinking processes in connection with temperature- and acidity changes. In winter (summer) the particles are likely to grow (shrink) through adsorption (evaporation) of water molecules as the temperature and weight fraction of H_2SO_4 in the sulphuric acid (H_2SO_4-H_2O) binary aerosol decrease.

Chapter 6

Summary and Conclusions

Aiming to improve the current knowledge of aerosol properties, in particular under background conditions, we propose a new retrieval algorithm which is based on the Optimal Estimation approach. The Optimal Estimation (OE) approach is based on Bayesian statistics, which provide a formalism (Bayes' theorem) that translates uncertainty in measurement space into uncertainty in state space. It takes a comprehensive approach to the multi-solution retrieval problem by identifying all possible solutions within a given measurement uncertainty and by characterizing these by means of probability density functions (pdf). Prior knowledge about aerosol properties at background levels is incorporated in the form of an a priori mean state and covariance matrix which define the probability density function of the physically sensible solution space, and by using the lognormal particle size distribution function to describe aerosol particle populations. The Bayesian solution to the retrieval problem is the probability density function of all possible states given a particular set of measurements. The mean and width of this pdf form the Optimal Estimation solution and the associated uncertainty estimate. The new OE algorithm is set up to retrieve the number density, median particle radius and distribution width of monomodal aerosol distributions from multi-spectral aerosol extinction data (as measured by the SAGE II satellite instrument) at four wavelengths in the ultraviolet, visible and near infrared part of the spectrum (at 0.386, 0.452, 0.525, and 1.020 μm).

In order to assess the performance of the new retrieval algorithm synthetic extinction coefficients were calculated for a 75% (by weight) sulphuric acid solution at 300 K (refractive indices by Palmer and Williams, 1975) at four spectral wavelengths (0.385, 0.452, 0.525 and 1.020 μm) and based on the 264 monomodal aerosol size distributions used to form the a priori pdf. Two test beds were generated by adding the following noise components: (a) the Minimum Noise Scenario (minNS) is characterised by a 1% Gaussian distributed random noise component added onto each of the four spectral extinction data; (b) the Maximum Noise Scenario (maxNS) is characterised by [60, 45, 30, 25]% Gaussian distributed random noise on the respective spectral channels [0.385, 0.452, 0.525,

1.020] μm. These numbers describe the typical range of experimental SAGE II extinction uncertainties. The majority of all measured (SAGE II) data can therefore be expected to be between these two extremes. In order to discard spurious retrieval solutions an ad hoc quality filter was developed based on several retrieval diagnostics. This filter achieves a good balance between maximizing the correlation between the 'good' (retrieved) and the correct solutions and minimizing data loss through rejection. In the case of both synthetic data sets approximately 88% of all retrieval solutions pass the screening.

6.1 Main Results

- Aerosol properties retrieved from synthetic extinction data were shown to be well correlated with the true solutions in both the minimum and the maximum noise scenarios.

- The coefficients describing the linear correlation between the retrieved and the correct aerosol properties are generally higher for the integrated aerosol properties (surface area density (A), volume density (V), and effective radius (R_{eff})) than for the size distribution parameters (number density (N), median particle radius (R), and distribution width (S)). This indicates that it is harder to correctly retrieve the lower order moments than the higher order moments of the size distribution, which is likely to be a consequence of the particle number density and size being more sensitive to the smallest particles (to which aerosol extinction measurements at visible wavelengths are insensitive) than surface area density or volume density (see Sec. 1.2).

- The retrieved uncertainties were found to be a good estimate of the true errors. Additional uncertainties due to forward model error, forward model parameter error or bimodal errors tend to be negligible.

- With little measurement noise (minNS), the retrieved uncertainties are observed to be of the order of 62% (σ_{N}), 24% (σ_{R}), 14% (σ_{S}), 22% (σ_{A}), and 11% (σ_{V}, σ_{Reff}). With large measurement noise (maxNS), the retrieved uncertainties are observed to be of the order of 75% (σ_{N}), 37% (σ_{R}), 26% (σ_{S}), 45% (σ_{A}), 34% (σ_{V}), and 15% (σ_{Reff}).

The new OE retrieval algorithm was then applied to SAGE II data measured during 6 months in 1999 (Sec. 5.1). These measurements are independent of the *a priori* data used in the retrieval algorithm. The results are compared to correlative *in situ* data and to aerosol properties retrieved by other researchers using different retrieval methods. We found:

- The OE number densities are observed to be larger and the median radii smaller and hence more realistic than results reported by [Bingen *et al., 2004b*] that were also retrieved from SAGE II data but using a regularized inversion technique. The retrieved OE errors are considerably smaller than those associated with the results by Bingen *et al.* (2004b).
- The OE surface area estimates are found be larger by about 20-50% than the PCA results (retrieved from the same SAGE II measurements) which are known to underestimate correlative in situ data by about 40% [Deshler *et al., 2003*].
- The OE volume densities are observed to be about 10 to 40% larger than the PCA volumes which tend to underestimate coincident in situ data by an estimated 30% [Steele *et al.*, 1999].
- The retrieved OE uncertainties are of the order of 69% (σ_{N}), 33% (σ_{R}), 14% (σ_{S}), 23% (σ_{A}), and 12% (σ_{V}), and 13% (σ_{Reff}).
- Compared to retrieval errors reported by other researchers the OE uncertainties are improved (σ_{N}, σ_{R}, σ_{S}) or of the same order of magnitude (σ_{A}, σ_{V}, σ_{Reff}).

The last part of Chapter 5 presents an example application, in which the new OE retrieval results are used to study seasonal differences in the retrieved aerosol properties on a global scale. In general the observed seasonal changes tend to be similar to seasonal differences observed by other researchers.

6.2 Conclusions

Based on these results we conclude:

- The new Optimal Estimation retrieval algorithm is able to successfully retrieve aerosol microphysical properties from synthetic extinction data that were generated from small aerosol particle ensembles typically observed under background aerosol conditions.
- Aerosol microphysical properties retrieved from measured SAGE II extinction data (December 1999) are closer to coincident in situ measurements and therefore likely to be closer to the correct solutions than previous results obtained through other retrieval methods.
- This indicates that the new OE retrieval algorithm is a successful approach to the aerosol retrieval problem and adds to the current knowledge by improving current estimates of aerosol properties in the lower stratosphere under low aerosol loading conditions.

6.3 Suggestions for Future Work

The aerosol properties as retrieved in 1999 can be combined with the current *a priori* data (which represents the aerosol background near 41°N) in order to generate a new *a priori* mean and standard deviation, that may be more representative of the global aerosol background conditions than the current *a priori* data and can therefore be expected to improve future retrieval results.

An altitude dependent first guess could be generated and implemented in the retrieval algorithm. We would expect to obtain improved retrieval results above all near the upper and lower ends of the vertical profiles of aerosol properties, where the greatest discrepancies between the retrieved and the correct (validation case, Sec. 4.5) or correlative *in situ* data (Sec. 5.2) were observed.

Bimodal Errors, that would arise in the presence of volcanically enhanced SAGE II extinction data, can be assessed and quantified.

The SAGE II climatology of aerosol properties generated by Dr. Steven Marsh using the new OE aerosol retrieval algorithm can be used to generate climatological monthly means. The climatological averages can be expected to differ somewhat from the monthly means of 1999 presented in this work because climatological data (a) remove interannual variability, and (b) can be expected to be free of any spatio-temporal bias (arising from uneven spatial and temporal data coverage) that may affect the monthly means of individual years.
The SAGE II climatology of OE aerosol properties can be downloaded from
http://www.atm.ox.ac.uk/project/PARTS/.

The retrieval algorithm can be adapted to other solar occultation instruments, for instance to SAGE III. As the SAGE III instrument measures on a wider spectral range than the SAGE II instrument (three additional aerosol channels near 0.76, 0.94 and 1.55 μm), it should be possible to retrieve bimodal aerosol size distribution parameters. This would mean that the new OE algorithm would be able to analyse volcanic aerosol as well.

The new Optimal Estimation retrieval algorithm can also be adapted to satellite instruments that operate in the infrared. Since the sensitivity to particle size is known to decrease with increasing particle size and/or increasing wavelength, it may not be possible to determine the size distribution parameters from IR measurements of aerosol extinction. In that case, the state vector would have to be changed to contain variables to which the IR measurements are more sensitive, such as surface area density and volume density.

Appendix A

The χ^2 Test

The χ^2 test is a way of testing whether a particular random vector belongs to a given distribution. If a vector z is a member of a Gaussian ensemble with zero mean and covariance S_z, then the quantity considered is $\chi^2 = z^T S_z^{-1} z$, which is twice the exponent in the Gaussian distribution. The χ^2 test asks the question: 'what fraction f of members of the Gaussian distribution have a probability density less than that of the vector being tested?'. If the fraction, f, is small, meaning that the probability of other χ^2 values being higher than χ_0^2, $P(\chi^2 \geq \chi_0^2)$ is small, then the vector is called an outlier at the $f \cdot 100\%$ level. If $P(\chi^2 \geq \chi_0^2)$ is less than 5%, the disagreement is called 'significant', if it is less than 1%, the disagreement is called 'highly significant' [Taylor, 1939]. The test is usually used to identify cases where χ^2 is too large, but it can also be used to detect cases where χ^2 is too small. In either case a significant result indicates that it is very likely that z is not part of the Gaussian distribution but rather an outlier.
When testing a variable z it is important that the covariance matrix S_z^{-1} is non-singular. This can be checked by looking at the eigenvector decomposition (for square matrices) or singular value decomposition of the covariance matrix S_z. In terms of the eigenvectors, L, and the matrix Λ of eigenvalues, λ_i, χ^2 can be expressed as

$$\chi^2 \;=\; z^T S_z^{-1} z \;=\; z^T L \Lambda^{-1} L^T z \;=\; z'^T \Lambda^{-1} z' \;=\; \sum_{i=1}^{p} \frac{z_i'^2}{\lambda_i} \tag{A.1}$$

χ^2 is the sum of separate terms for p independent normally-distributed random variables $z' = L^T z$ with each z_i' having a corresponding variance λ_i. This means that in the case of a singular covariance matrix where at least one of the eigenvalues is zero, the above expression cannot readily be used. Instead, a pseudo-inverse $S_z'^{-1} = L' \Lambda'^{-1} L'^T$ has to be constructed, in which only the p non-zero eigenvalues and corresponding eigenvectors are retained in Λ' and L', respectively.
χ^2 has p degrees of freedom, an expected value of $\varepsilon\{\chi^2\} = p$ and a variance $var(\chi^2) = 2p$. For large p it tends to a normal distribution.

Appendix B

χ^2 test for $[\hat{y} - y]$

A χ^2 significance test performed on the difference $\delta y = (F(\hat{x}) - y) = (\hat{y} - y)$ between the true or measured extinction and the retrieved extinction provides a diagnostic for the quality of a retrieved vector $\hat{x}$. In the trust region of a retrieved solution, where the approximation $K \approx K|_{\hat{x}}$ holds true, $\delta y = (\hat{y} - y)$ can be expressed in terms of the linearised forward model

$$\begin{aligned} \hat{y} - y &= K(\hat{x} - x) - \epsilon_y \\ &= K[(GK - I_n)(x - x_a) + G\epsilon_y] - \epsilon_y \\ &= (KG - I_m)[K(x - x_a) + \epsilon_y]. \end{aligned} \tag{B.1}$$

On substituting for G (Eq. 4.6) and applying the equality of $(K^T S_\epsilon^{-1} K + S_a^{-1})^{-1} = S_a K^T (K S_a K^T + S_\epsilon)$ the covariance of $(\hat{y} - y)$ becomes

$$\begin{aligned} S_{(\hat{y}-y)} &= \varepsilon\{(F(\hat{x}) - y)(F(\hat{x}) - y)^T\} \\ &= (KG - I_m)(K S_a K^T + S_{\epsilon_y})(KG - I_m)^T \\ &= S_\epsilon (K S_a K^T + S_\epsilon)^{-1} S_\epsilon \end{aligned} \tag{B.2}$$

Since S_ϵ is off full rank, m, and therefore $S_{(F(\hat{x})-y)}$ is as well, the χ^2 test is:

$$\chi^2 = (\hat{y} - y) S_{(\hat{y}-y)}^{-1} (\hat{y}) - y)^T \tag{B.3}$$

Bibliography

Anderson, J., Brogniez, C., Cazier, L., Saxena, V. K., Lenoble, J. and McCormick, M. (2000). Characterization of aerosols from simulated SAGE III measurements applying two retrieval techniques. *J. Geophys. Res.*, **105** (D2), 2013–2027.

Anderson, J. and Saxena, V. K. (1996). Temporal changes of Mount Pinatubo aerosol characteristics over northern midlatitudes derived from SAGE II extinction measurements. *J. Geophys. Res.*, **101** (D14), 19,455–19,463.

Atkins, P. W. (1997). *Physical chemistry.* fifth edition 1997. Aufl., Oxford University Press.

Biermann, U. M. *et al.* (2000). Absorption spectra and optical constants of binary and ternary solutions of H_2SO_4, HNO_3 and H_2O in the mid infrared at atmosperic temperatures. *J. Phys. Chem. A*, **104** (4), 783–793.

Bingen, C., Fussen, D. and Vanhellemont, F. (2004a). A global climatology of stratospheric aerosol size distribution parameters derived from SAGE II data over the period 1984-2000: 1. methodology and climatological observations. *J. Geophys. Res.*, **109** (D06201), doi:10.1029/2003JD003518.

Bingen, C., Fussen, D. and Vanhellemont, F. (2004b). A global climatology of stratospheric aerosol size distribution parameters derived from SAGE II data over the period 1984-2000: 2. reference data. *J. Geophys. Res.*, **109** (D06202), doi:10.1029/2003JD003511.

Bingen, C., Vanhellemont, F. and Fussen, D. (2002). A new regularized inversion method for the retrieval of stratospheric aerosol size distribution applied to 16 year SAGE II data (1984-2000): Method, results, and validation. *Ann. Geophys.*, **21**, 797–804.

Bluth, G. J. S. *et al.* (1992). Global tracking of the SO_2 clouds from the June 1991 Mount Pinatubo eruptions. *Geophys. Res. Lett.*, **19**, 151–154.

Chen, Y. *et al.* (1998). Single particle analyses of ice nucleating aerosols in the upper troposphere and lower stratosphere. *Geophys. Res. Lett.*, **25**, 1391–1394.

Chu, W. P. (1986). Inversion of SAGE II measurements. In *Sixth Conference on Atmospheric Radiation, AMS*, Williamsburg, VA. Am. Meteor. Soc.

Chu, W. P. and McCormick, M. P. (1979). Inversion of stratospheric aerosol and gaseous constituents from spacecraft solar extinction data in the 0.38-1.0 μm wavelength region. *Appl. Opt.*, **18**, 1404–1413.

Chu, W. P., McCormick, M. P., Lenoble, J., Brogniez, C. and Pruvost, P. (1989). SAGE II inversion algorithm. *J. Geophys. Res.*, **94** (D6), 8339–8351.

Colbeck, I. (1998). *Physical and chemical proprties of aerosols*. Blackie Academic & Professional.

Deepak, A. and Box, G. P. (1982). *Atmospheric aerosols: their formation, optical properties, and effects*. Spectrum Press, Hampton, Virginia.

Deshler, T., Hervig, M. E., Hofmann, D. J., Rosen, J. M. and Liley, J. B. (2003). Thirty years of *in situ* stratospheric aerosol size distribution measurements from Laramie, Wyoming (41°N) using balloon-borne instruments. *J. Geophys. Res.*, **108** (D5), doi:10.1029/2002JD002514.

Deshler, T., Hofmann, D. J., Johnson, B. J. and Rozier, W. R. (1992). Balloonborne measurements of the Pinatubo aerosol size distribution and volatility at Laramie, Wyoming during the summer of 1991. *Geophys. Res. Lett.*, **19**, 199–202.

Deshler, T., Johnson, V. J. and Rozier, W. R. (1993). Balloonborne measurements of Pinatubo aerosol during 1991 and 1992 at 41°N: Vertical profiles, size distributions, and volatility. *Geo. Res. Lett.*, **20** (D14), 1435–1438.

Franssens, G. (1999). Spectral interpolation and inversion of aerosol extinction coefficients. *J. Aerosol Sci.*, **30**, S565–S566. Suppl. 1.

Fu, W. *et al.* (1999). Modeling of scattering and absorption by nonspherical cirrus ice particles at thermal infrared wavelengths. *J. Atmos. Sci.*, **56**, 2937–2945.

Goodman, J. *et al.* (1994). Evolution of Pinatubo aerosol near 19 km altitude over western North America. *Geophys. Res. Lett.*, **21**, 1129–1132.

Graf, H.-F. *et al.* (1993). Pinatubo eruption winter climate effects: Model versus observations. *Clim. Dyn.*, **9**, 81–93.

Grainger, R. G. (1990). *The calculation of cloud parameters from AVHRR data*. Dissertation, University of Auckland, New Zealand.

Grainger, R. G. *et al.* (1993). Infrared absorption by volcanic stratospheric aerosols observed by ISAMS. *Geophys. Res. Lett.*, **20** (12), 1283–1286.

Grainger, R. G. *et al.* (1995). Stratospheric aerosol effective radius, surface, and volume estimated from infrared measurements. *J. Geophys. Res.*, **100**, 16,507–16,518.

Granier, C. and Brasseur, G. (1992). Impact of heterogeneous chemistry on model predictions of ozone changes. *J. Geophys. Res.*, **97** (D16), 18,015–18,033.

Grant, W. B., Browell, E. V., Long, C. S., Stowe, L. L., Grainger, R. G. and Lambert, A. (1996). Use of aerosols resulting from volcanic eruptions to study the tropical stratospheric reservoir, its boundary, and transport to northern latitudes. *J. Geophys. Res.*, **101**, 3973–3988.

Heintzenberg, J., Muller, H., Quenzel, H. and Thomalla, E. (1981). Information content of optical data with respect to aerosol properites: Numerical studies with a randomized minimization-search-technique inversion algorithm. *Appl. Opt.*, **20**, 1308–1315.

Heintzenberg, J. *et al.* (1996). On the composition of non-volatile material in upper tropospheric aerosols and cirrus crystals. *Atmos. Res.*, **41**, 81–88.

Hervig, M. E., Deshler, T. and Russell III, J. M. (1998). Aerosol size distributions obtained from HALOE spectral extinction measurements. *J. Geophys. Res.*, **103** (D1), 1573–1583.

Hinds, W. C. (1998). *Aerosol Technology - Properties, behaviour, and measurments of airborne particles.* John Wiley & sons.

Hofmann, D. J. (1990). Increase in the stratospheric background sulfuric acid aerosol mass in the past 10 years. *Science*, **248**, 996–1000.

Hofmann, D. J. (1991). Aircraft sulphur emissions. *Nature*, **349**, 659.

Hofmann, D. J., Carroll, D. E. and Rosen, J. M. (1975). Estimate of the contribution of the space shuttle effluent to the natural stratospheric aerosol. *Geophys. Res. Lett.*, **2**, 113–116.

Hofmann, D. J. and Solomon, S. (1989). Ozone destruction through heterogeneous chemistry following the eruption of El Chichón. *J. Geophys. Res.*, **94** (D4), 5029–5041.

Holton, J. R. *et al.* (1995). Stratosphere-troposphere exchange. *Reviews of Geophysics*, **33** (4), 403–439.

Junge, C. and Manson, J. (1961). Stratospheric aerosol studies. *J. Geophys. Res.*, **66** (7), 2163–2183.

Kärcher, B. and Solomon, S. (1999). On the composition and optical extinction of particles in the tropopause region. *J. Geophys. Res.*, **104** (D22), 27,441–27,459.

Kent, G. S., Winker, D. M., Osborn, M. T. and Skeens, K. M. (1993). A model for the separation of cloud and aerosol in SAGE II occultation data. *J. Geophys. Res.*, **98**, 20,725–20,735.

Kiehl, J. T. *et al.* (2000). Radiative forcing due to sulfate aerosols from simulations with the National Center for Atmospheric Research community climate model, version 3. *J. Geophys. Res.*, **105** (D1), 1441–1457.

Kodera, K. (1994). Influence of volcanic eruptions on the troposphere through stratospheric dynamical processes in the northern hemisphere winter. *J. Geophys. Res.*, **21**, 1273–1282.

Labitzke, K. G. and van Loon, H. (1999). *The stratosphere – phenomena, history and relevance*. Springer.

Lacis, A., Hansen, J. and Sato, M. (1992). Climate forcing by stratospheric aerosols. *Geophys. Res. Lett.*, **19**, 1607–1610.

Lambert, A., Grainger, R. G., Remedios, J. J., Reburn, W. J., Rodgers, C. D., Taylor, F. W., Roche, A. E., Kumer, J. B., Massie, S. T. and Deshler, T. (1996). Validation of aerosol measurements from the Improved Stratospheric And Mesospheric Sounder. *J. Geophys. Res.*, **101** (D6), 9811–9830.

Lambert, A., Grainger, R. G., Rodgers, C. D., Taylor, F. W., Mergenthaler, J. L., Kumer, J. B. and Massie, S. T. (1997). Global evolution of the Mt. Pinatubo volcanic aerosols observed by the infrared limb-sounding instruments CLAES and ISAMS on the Upper Atmosphere Research Satellite. *J. Geophys. Res.*, **102** (D1), 1495–1512.

Lin, N.-H. and Saxena, V. K. (1992). Characteristics of Antarctic stratospheric aerosols during the 1987 ozone depletion episode based on SAGE II satellite observations. *J. Atmos. Sci.*, **57**, 442–451.

Lu, C.-H. *et al.* (2003). Retrieval analysis of aerosol integral properties from simulated extinction at SAGE II and HALOE wavelengths. *J. Geophys. Res.*, **108** (D7), 1–9.

Lu, J., , Mohnen, V. A., Yue, G. K. and Jäger, H. (1997). Intercomparison of multiplatform stratospheric aerosol and ozone observations. *J. Geophys. Res.*, **102** (D13), 16,127–16,136.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, **11**, 431–441.

Massie, S. T. *et al.* (1996b). Evolution of the infrared properties of the Mount Pinatubo aerosol cloud over Laramie, Wyoming. *J. Geophys. Res.*, **101** (D17), 23,007–23,019.

McCormick, M. P. (1987). SAGE II: an overview. *Adv. Space Res.*, **7** (3), 219–226.

McCormick, M. P., Thomason, L. W. and Trepte, C. R. (1995). Atmospheric effects of the Mt. Pinatubo eruption. *Nature*, **373**, 399–403.

McMaster, L. R. (1986). Stratospheric aerosol and gas experiment SAGE II. Williamsburh, VA. Am. Meteorol. Soc. paper presented at Sixth Conference on Atmospheric Radiation.

Mie, G. (1908). Beitraege zur Optik trueber Medien speziell kolloidaler Metalloesungen. *Annals of Physics*, **25**, 377.

Molina, M. J. *et al.* (1993). Physical chemistry of the $H_2SO_4/HNO_3/H_2O$ system: implications for polar stratospheric clouds. *Science*, **261**, 1418–1423.

Murphy, D. M. *et al.* (1998). Organics, meteorite material, mercury, and other elements in high altitude aerosols. *Science*, **282**, 1664–1669.

Nair, P. V. N. and Vohra, K. G. (1975). Growth of aqueous sulphuric acid droplets as a function of relative humidity. *J. Atmos. Sci.*, **6**, 265–271.

Palmer, K. F. and Williams, D. (1975). Optical constants of sulfuric acid: application to the clouds of Venus? *Appl. Opt.*, **14** (1), 208–219.

Pinkley, L. W. and Williams, D. (1976). The infrared optical constants of sulfuric acid at 250 K. *J. Opt. Soc. Am.*, **66** (2), 122–124.

Pollack, J. B. *et al.* (1976). Estimates of the climatic impact of aerosols produced by space shuttles, SST's, and other high flying aircraft. *J. Appl. Meteorol.*, **15**, 247–258.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical recipes in FORTRAN 77 - The art of scientific computing.* 2nd ed.. Aufl., Cambridge University Press, New York.

Pueschel, R. F. (1996). Stratospheric aerosols: Formation, properties, effect. *J. Aerosol Sci.*, **27**, 383–402.

Quenzel, H. (1970). Determination of size distribution of atmospheric aerosol particles from spectral solar radiation measurements. *J. Geophys. Res*, **75**, 2915–2921.

Robock, A. and Mao, J. (1995). The volcanic signal in surface temperature observations. *J. Clim.*, **8**, 1086–1103.

Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding – Theory and practice*, Bd. 2 von *Atmospheric, Oceanic and Planetary Physics*. World Scientific, Singapore.

Rosen, J. M. (1991). A comparison of measured and calculated optical properties of atmospheric aerosols at infrared wavelengths. *J. Geophys. Res.*, **96**, 5229–5235.

Russel, B. P. *et al.* (1981). Satellite and correlative measurements of the stratospheric aerosol: I– An optical model for data conversions. *J. Atmos. Sci.*, **38**, 1279.

Russell, P. B. and Hamill, P. (1984). Spatial variation of stratospheric aerosol acidity and model refractive index: Implications of recent results. *J. Atmos. Sci.*, **42**, 1781–1790.

Russell, P. B. *et al.* (1996). Global to microscale evolution of the Pinatubo volcanic aerosol, derived from diverse measurements and analyses. *J. Geophys. Res.*, **101**, 18,745–18,764.

Saxena, V. K., Anderson, J. and Lin, N.-H. (1995). Changes in antarctic stratospheric aerosol characteristics due to volcanic eruptions as monitored by SAGE II satellite. *J. Geophys. Res.*, **100** (D6), 16,735–16,751.

Seinfeld, J. H. and Pandis, S. N. (1998). *Atmospheric chemistry and physics – From air pollution to climate change*. Wiley & Sons, Inc., New York.

Semmler, M., Luo, B. and Koop, T. (2003). Properties of ternary $H_2SO_4 - (NH_4)_2SO_4 - H_2O$ at UT/LS conditions. In *EGS - AGU - EUG Joint Assembly*, Nice, France. EGU. Abstract no. 8691.

Shen, T.-L. *et al.* (1995). *Stratospheric pollution and ozone depletion*, S. 394–442. Composition, Chemistry, and Climate of the Atmosphere. Van Nostrand Reinhold, New York.

Solomon, S. *et al.* (1986). On the depletion of Antarctic ozone. *Nature*, **321**, 755.

SPARC, S. (2006). *Assessment of Stratospheric Aerosol Properties (ASAP)*. Stratospheric Processes and Their Role in Climate (SPARC), A project of the WMO/ICSU/IOC World Climate Research Programme, WCRP-124, WMO/TD- No. 1295, SPARC Report No. 4.

Steele, H. M. and Hamill, P. (1981). Effects of temperature and humidity on the growth and optical properties of sulphuric acid-water droplets in the stratosphere. *J. Aerosol Sci.*, **12**, 517–528.

Steele, H. M., Lumpe, J. D., Turco, R. P., Bevilacqua, R. M. and Massie, S. T. (1999). Retrieval of aerosol surface area and volume densities from extinction measurements: Application to POAM II and SAGE II. *J. Geophys. Res.*, **104** (D8), 9325–9336.

Steele, H. M. and Turco, R. P. (1997). Retrieval of aerosol size distributions from satellite extinction spectra using constrained linear inversion. *J. Geophys. Res.*, **102** (D14), 16,737–16,747.

Talbot, R. W. *et al.* (1998). Influences of vertical transport on free tropospheric aerosols over the central USA in springtime. *Geophys. Res. Lett.*, **25**, 1367–1370.

Taylor, J. R. (1939). *An introduction to Error Analysis.* University Science Books, Sausalito, CA. Second Edition.

Thomason, L. W. (1991). A diagnostic stratospheric aerosol size distribution inferred from SAGE II measurements. *J. Geophys. Res.*, **96** (D12), 22,501–22,508.

Thomason, L. W. and Burton, S. P. (2006). SAGE II measurements of stratospheric aerosol properties at non-volcanic levels. *ACP.* submitted to ACP, 2006.

Thomason, L. W., Kent, G. S., Trepte, C. R. and Poole, L. R. (1997a). A comparison of the stratospheric aerosol background periods of 1979 and 1989-1991. *J. Geophys. Res.*, **102** (D3), 3611–3616.

Thomason, L. W. and Osborn, M. T. (1992). Lidar conversion parameters derived from SAGE II extinction measurements. *Geophys. Res. Lett.*, **19**, 1655–1658.

Thomason, L. W. and Poole, L. R. (1993). Use of stratospheric aerosol properties as diagnostics of Antarctic vortex processes. *J. Geophys. Res.*, **98**, 23,002–23,012.

Thomason, L. W., Poole, L. R. and Deshler, T. (1997b). A global climatology of stratospheric aerosol surface aerea density deduced from Stratospheric Aerosol and Gas Experiment II measurements: 1984-1994. *J. Geophys. Res.*, **102** (D7), 8967–8976.

Timmreck, C., Graf, H.-F. and Kirchner, I. (1999). A one and half year interactive MA/ECHAM4 simulation of Mount Pinatubo aerosol. *J. Geophys. Res.*, **104**, 9337–9359.

Tolbert, M. A. (1994). Sulfate aerosols and stratospheric cloud formation. *Science*, **264**, 527–528.

Torres, O. *et al.* (1998). Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: Theoretical basis. *J. Geophys. Res.*, **103** (D14), 17,099–17,110.

Turco, R. P. *et al.* (1980). Stratospheric aerosol modification by supersonic transport and space shuttle operations - climate implications. *J. Appl. Meteorol.*, **19**, 78–89.

Turco, R. P. *et al.* (1980a). OCS, stratospheric aerosols and climate. *Nature*, **283**, 283–286.

Twomey, S. (1974). Information content in remote sensing. *Appl. Opt.*, **13**, 942–945.

Twomey, S. (1977). *Introduction to the mathematics of inversion in remote sensing and indirect measurements*. Dover Publications, Inc., Mineola, New York.

van de Hulst, H. C. (1957). *Light scattering by small particles*. Wiley & Sons, New York.

Viera, G. and Box, M. A. (1985). Information content analysis of aerosol remote-sensing experiments using an analytic eigenfunction theory: anomalous diffraction approximation. *Appl. Opt.*, **24** (24), 4525–4533.

Wang, P. H., Kent, G. S., McCormick, M. P., Thomason, L. W. and Yue, G. K. (1996). Retrieval analysis of aerosol size distribution with simulated extinction measurements at SAGE III wavelengths. *Appl Opt.*, **35** (3), 433–440.

Wang, P. H., McCormick, M. P., Swissler, T. J., Osborn, M. T., Fuller, W. J. and Yue, G. K. (1989). Inference of stratospheric aerosol composition and size distribution from SAGE II satellite measurements. *J. Geophys. Res.*, **94**, 8435–8446.

Yue, G. K. (1999). A new approach to retrieval of aerosol size distributions and integral properties from SAGE II aerosol extinction spectra. *J. Geophys. Res.*, **104** (D22), 27,491–27,506.

Yue, G. K. and Deepak, A. (1983). Retrieval of stratospheric aerosol size distribution from atmospheric extinction of solar radiation at two wavelengths. *Appl. Opt.*, **22**, 1639–1645.

Yue, G. K., Lu, J., Mohnen, V. A., Wang, P.-H., Saxena, V. K. and Anderson, J. (1997). Retrieving aerosol optical properties from moments of the particle size distribution. *Geophys. Res. Lett.*, **24** (6), 651–654.

Yue, G. K., McCormick, M. and Chu, W. (1986). Retrieval of composition and size distribution of stratospheric aerosols with the SAGE II satellite experiment. *J. Atmos. Oceanic Technol.*, **3**, 371–380.

Yue, G. K., Poole, L. R., Wang, P.-H. and Chiou, E. W. (1994). Stratospheric aerosol acidity, density, and refractive index deduced from SAGE II and NMC temperature data. *J. Geophys. Res.*, **99** (D2), 3727–3738.

Yue, G. K., Thomason, L. W., Poole, L. R., Wang, P.-H., Baumgardner, D. and Dye, J. E. (1995). Aerosol surface areas deduced from early 1993 SAGE II data and comparisons with stratospheric photochemistry, aerosols, and dynamics expedition measurements. *Geophys. Res. Lett.*, **22**, 2933–2936.

Yue, G. K. *et al.* (1989). Comparative studies of aerosol extinction measurements made by the SAM II and SAGE II satellite experiments. *J. Geophys. Res.*, **94** (D6), 8412–8424.

ACKNOLEDGEMENTS

This research could not have been undertaken and this thesis not completed without the support of many. And so I would like to acknowledge and give my sincere thanks to

- my supervisors Dr Adrian McDonald, Dr Don Grainger and Dr Darlene Heuff for their advice, feedback and support during my thesis work;
- NASA Langley Research Center (LaRC) for providing the SAGE II aerosol extinction data, and Dr Larry Thomason (NASA LaRC) for information concerning the PCA retrieval approach;
- Prof Terry Deshler, from the Department of Atmospheric Science at the University of Wyoming, for providing the *in situ* measurements;
- Dr Don Grainger and the Department of Atmospheric, Oceanic and Planetary Physics (AOPP) at the University of Oxford for hosting me during an extended research stay. Most helpful were the discussions with Dr Clive Rogers, Dr Anu Dudhia and the students of the atmospheric group;
- my fellow students and the staff of the atmospheric group for fruitful discussion and feedback;
- the people who provided IT support for the retrieval runs, in particular Orlon Petterson, Graham Planck and Paul Arnold;
- my mentors Dr Jenny Adams and Dr Simon Brown for their advice and support;
- all members of the Department of Physics and Astronomy of the University of Canterbury for providing such a pleasant work environment;
- the Department of Physics and Astronomy for providing funding which enabled me to travel to Oxford (UK) for the research stay at AOPP at the University of Oxford;
- the University of Canterbury, which provided the funding for the PhD project itself and provided me with a Doctoral Scholarship;
- my family (in particular my husband Andreas and son Liam) for providing continuing support in particular during the final stages of this PhD project.